Investigating the structure and architecture of the NuA4 subcomplex TINTIN

Investigating the Structure and Architecture of the NuA4 Subcomplex TINTIN

Emir Acıyan

PhD

Birkbeck, University of London

0.1 Declarations

The work presented within this thesis is my own and contains nothing that is the outcome of collaboration, except where specifically mentioned in the text (an accompanying statement of contribution is submitted with this thesis).

This PhD thesis has been prepared in accordance with all Birkbeck, University of London regulations.

Emir Acıyan 26 September, 2019

0.2 Acknowledgements

Thanks are extended to my supervisors Dr Alan Cheung and Dr Konstantinos Thalassinos, their advice and insights have been abundant and extremely useful when addressing the challenges that came with this project.

Thanks to Dr Cara Vaughan for her advice as my thesis chair, Dr Nikos Pinotsis, whose expertise in SAXS was invaluable as I carried out data collection and processing. Charles Eldrid of the Thalassinos lab must also be thanked, who was kind enough to help collect native mass spectrometry data and carried out analysis of the raw data. Thanks to the MRC for the funding of this project and the ISMB (UCL and Birkbeck) for giving me the opportunity to carry out this research.

And thanks must also be given to my friends and colleagues at the ISMB, both past and present, who I had the privilege of working beside throughout the challenges of this project.

Finally I would like to thank my family for their support and encouragement, and for reminding me what really matters.

0.3 Abstract

The regulation of gene expression is essential for the correct function of virtually all cellular processes and misregulation has a range of pathological consequences, including Alzheimer's disease and cancer. The NuA4 coactivator complex is an essential eukaryotic histone acetyltransferase, that promotes transcriptional activation and DNA repair. The complex targets the N-terminal tail of H4 and H2A, recruited to specific loci by DNA-binding transcriptional activators. NuA4 is highly modular, with several subcomplexes that can exist independently, or shared with other coactivator complexes such as SAGA, allowing for targeted acetylation of specific gene loci.

An investigation of the structure and function of the NuA4 subcomplex TINTIN, is presented herein, a trimeric module comprised of the subunits Eaf3, Eaf5 and Eaf7. TINTIN has been implicated in promoting transcription elongation, via a proposed interaction with RNA polymerase II and regulating nucleosome stability. TINTIN has been recombinantly expressed, purified and subjected to crystallography trials, which failed, despite multiple truncated TINTIN constructs that produced complexes of higher stability and purity. Nevertheless, these constructs provided insights into TINTIN architecture and identified key determinants for complex stability and major differences with the orthologous complex found in humans. A combination of SEC-MALS, native mass spectrometry and small-angle X-ray scattering analysis, provide low-resolution information, regarding the overall topology and oligomeric state of the complex.

TINTIN exists predominantly in its heterotrimeric state, though the interaction between Eaf5 and the rest of the complex is relatively labile. The complex adopts an extended conformation, with some indications of partial flexibility/disorder. Additional factors may be required to stabilise TINTIN, such as subunits of the parent NuA4 complex and/or the RNA Pol II CTD, which pull-down experiments presented herein, suggest directly interacts with the complex.

0.3.1 Brief Aims and Objectives

The TINTIN subcomplex was originally identified by immunoprecipitation from yeast (Rossetto et al. 2014), which showed that a complex of NuA4 subunits Eaf3, Eaf5 and Eaf7, could be purified from cellular extracts without its NuA4 parent complex. Furthermore, TINTIN localised to gene bodies and correlated highly with the presence of H3K36me3 and Pol II CTD phosphorylation at Ser2. However, the molecular features of TINTIN structure and its putative association with modified histones and Pol II were unknown. The aim of this thesis was to explore these aspects of yeast TINTIN function, by structural and biochemical analysis, with the following objectives:

1. To express, purify and structurally characterise the TINTIN complex, using a combination of X-ray crystallography, native ion mobility-mass spectrometry and SAXS.

2. To generate a series of truncated TINTIN constructs, followed by expressions/purifications, to identify essential components for complex assembly and function.

3. To explore the relationship of TINTIN to the orthologous complex MRG15/MRGBP, which lacks an ortholog of Eaf5.

4. To verify an interaction between TINTIN and Pol II CTD, using pull-down assays with purified TINTIN and phosphorylated peptides, that mimic the CTD at different stages of transcription.

5. To determine whether TINTIN could interact directly with methylated histone H3 and to compare this affinity, to that of the isolated chromodomain of Eaf3.

Contents

0.4 Abbreviations

International system of units (SI) abbreviations and standard notation for chemical elements/formulae are used in this thesis. Additional abbreviations include:

1. Act1: Actin-1.
2. AI media: Autoinduction media.
3. Å: Angstroms.
4. Arp4: Actin-related protein 4.
5. ATP: Adenosine triphosphate.
6. bp: base-pairs.
7. BSA: Bovine serum albumin.
8. CCS: Collision cross sections.
9. ChiP: Chromatin immunoprecipitation.
10. CID: Collision induced dissociation.
11. Cryo-EM: Cryo-electron microscopy.
12. CSD: Charge state distribution.
13. CTD: C-terminal domain.
14. Ctk1: Carboxy-terminal kinase.
15. CV: Column volumes.
16. D_{max}: Maximum dimension within a molecule.
17. DFL: Dual-function loop.

18. DMSO: Dimethyl sulfoxide.

19. DNA: Deoxyribonucleic acid.

20. DTT: Dithiothreitol.

21. Eaf1-7: Esa1 associated factors 1-7.

22. EDTA: Ethylenediaminetetraacetic acid.

23. Epl1: Enhancer of polycomb-like protein 1.

24. Esa1: histone acetyltransferase protein.

25. ESI: electrospray ionisation.

26. FACT: Facilitates chromatin transcription remodelling complex.

27. FATC: FAT(FRAP, ATM, TRRAP) C-terminal domain.

28. FL: full-length protein.

29. GST: Glutathione S-transferase.

30. H2A: Histone 2A.

31. H2B: Histone 2B.

32. H3: Histone 3.

33. H4: Histone 4.

34. HAT: Histone acetyltransferase.

35. HBL: Histone binding loop.

36. HDAC: Histone deacetylase.

37. HEAT: Huntingtin, elongation factor 3 (EF3), protein phosphatase 2A (PP2A), and the yeast kinase TOR1 domain.

38. HFD: Histone fold domain.

39. HPLC: High performance liquid chromatography.

40. HSA: Helicase /SANT-associated domain.

41. I(0): SAXS forward scattering intensity.

42. IEX: Ion-exchange chromatography.

43. IM-MS: Ion mobility-mass spectrometry.

44. ING1: Inhibitor of growth protein-1.

45. IPTG: Isopropyl β- d-1-thiogalactopyranoside.

46. kDa: kilo Daltons.

47. LB: Luria broth.

48. LS: Light-scattering signal

49. M/Z: Mass/Charge ratio.

50. MALDI: Matrix assisted laser desorption/ionisation.

51. MES: 2-(N-morpholino)ethanesulfonic acid buffer.

52. MMS: Methyl methanesulfonate.

53. MPA: Methiopropamine.

54. MRG15: MORF related gene 15 protein.

55. MRGBP: MRG15 binding protein.

56. mRNA: Messenger RNA.

57. MSMS: Tandem mass spectrometry.

58. MW: Molecular weight.

59. Native MS: Native mass spectrometry.

60. NCP: Nucleosome core particle.

61. NMR: Nuclear magnetic resonance.

62. NSD: Normalized spatial discrepancy.

63. NuA4: Nucleosome acetyltransferase of H4.

64. OD: Optical density.

65. D_{max}: Maximum dimension of a species (SAXS).

66. PCR: Polymerase chain reaction.

67. PDB: Protein data bank.

68. PEG: Polyethylene glycol.

69. PIC: Pre-initiation complex.

70. PIKK: Phosphatidylinositol 3-kinase-related kinase family.

71. Pol II: RNA polymerase II.

72. PTM: Post-translational modifications.

73. R_{free}: Crystallographic R-factor (from 5% of data).

74. R_g: Radius of gyration.

75. R_{work}: Measure of agreement between crystallography model and experimental data

76. rcf: Relative centrifugal source.

77. RI: Refractive index.

78. RNA: Ribonucleic acid.

79. Rpd3S: Rpd3 small complex.

80. rpm: Revolutions per minute.

81. SAGA: Spt-Ada-Gcn5-acetyltransferase complex.

82. SAXS: Small-angle X-ray scattering.

83. SDS-PAGE: Sodium dodecyl sulfate polyacrylamide gel electrophoresis.

84. SEC-MALS: Size-exclusion chromatography, multi-angle light scattering.

85. SPB: Spindle pole body.

86. Strep: Streptavidin.

87. SWI/SNF: SWItch/Sucrose Non-Fermentable.

88. SWR1: Chromatin remodelling complex.

89. T7: Bacteriophage T7 polymerase.

90. TAP: Tandem-affinity purification.

91. TB: Terrific broth.

92. TBP: TATA box binding protein.

93. TEEAA: Tra1, Eaf1, Eaf5, actin and Arp4.

94. TINTIN: Trimer independent of NuA4 involved in transcription interactions with nucleosomes.

95. Tip60: Human histone acetyltransferase of H4.

96. T_m: Temperature at which ~50% of protein has denatured in thermofluor assays.

97. TOR: Target of rapamycin.

98. Tra1: Transcription-associated protein.

99. TRRAP:Transformation/Transcription domain associated protein.

100. TS: Temperature sensitive.

101. UAS: Upstream activator sequence.

102. UV: Ultraviolet signal.

103. v/v: Volume/volume.

104. w/v: Weight/volume.

105. WT: Wild-type.

106. Yaf9: Yeast homolog of the human leukemogenic protein AF9.

107. Yng2: Yeast ING1 homolog.

List of Figures

List of Tables

Chapter 1

Introduction

1.1 What is Chromatin?

Eukaryotes, unlike their prokaryotic counterparts, have evolved to package their genetic material into large complexes of DNA and histone proteins, termed chromatin. The structure of chromatin has long been associated with the regulation of cellular processes and this includes, but is not limited to; transcription, DNA replication and cell cycle progression.

We have been aware of chromatin since the early 20^{th} century, by staining eukaryotic nuclei, which revealed condensed structures of DNA and protein, at various stages of mitosis (Huisinga et al. 2006). Chromatin arranges itself into regions of varying density, with the denser regions (heterochromatin) associated with reduced levels of gene expression, compared to the less compact euchromatin (Lorentz et al. 1992).

Chromatin is dynamic and can transition between these two states, depending on the stage of the cell cycle, or to allow for the activation of specific gene loci (Grewal and Moazed 2003, Lai and Pugh, 2017).

The basic repeating unit of chromatin is the nucleosome core particle (NCP), which has been structurally characterised to a high-resolution (Luger et al. 1997) (**Figure 1.1**). These NCPs assemble into a "beads on a string" arrangement, which consists of long fibres of DNA, punctuated by nucleosomes, made visible by Olins and Olins (1974) with negative staining.

Figure 1.2 presents a schematic of the repeating beads on a string array, which is generic for eukaryotes. NCPs are the target of large chromatin modifying machinery, that can result in subtle changes, such as the post-translational modification of specific histone residues, to the complete exchange/movement of histone subunits.

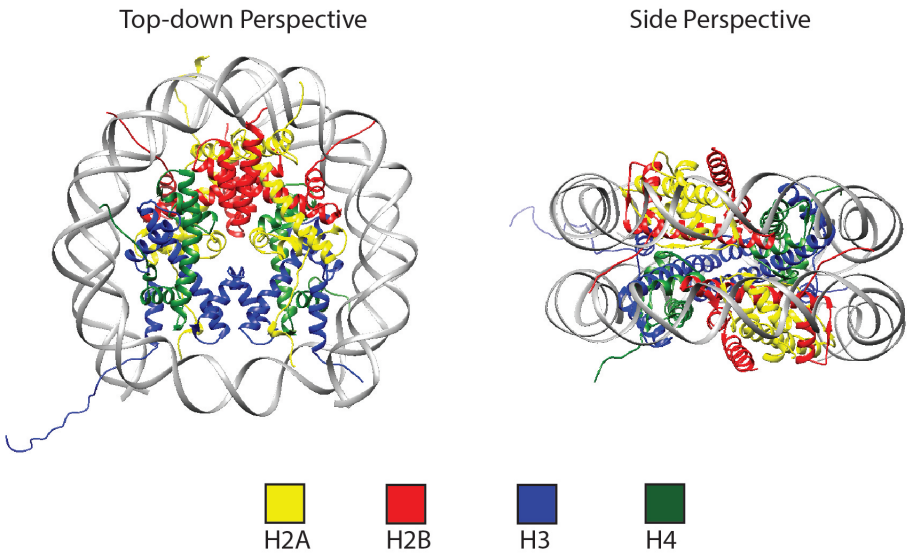

Figure 1.1: 2.8 Å resolution crystal structure of the human nucleosome core particle (Luger et al. 1997). 147 DNA base pairs (grey) wrap around a protein octamer, consisting of two copies of the four following proteins: H2A, H2B, H3 and H4 (which have been colour coded). Histones form histone fold domains (HFD), helical regions connected by unstructured loops, that arrange in a head to tail manner. The HFDs are essential for forming contacts between individual histones within the nucleosome, as well as contact with the surrounding DNA.

These varying modifications can have profound consequences on the fate of the cell and understanding the mechanisms by which chromatin is modified, has given birth to an entire field of molecular biology. Insights into the structural/chemical modification of chromatin, has helped better understand certain disease pathologies.
For example, gain of function mutations in the "guardian of the genome" p53, have been shown to lead to increased methylation and acetylation of chromatin, which in turn leads to a pathogenic increase in the expression of tumour promoting genes (Zhu et al. 2015).

The pathological mechanisms of other diseases, such as Alzheimer's, have also been implicated in the structure/state of chromatin (Frost et al. 2014). Furthermore, chromatin modification has also been shown to play a critical role in DNA damage repair pathways (Bird et al. 2002).

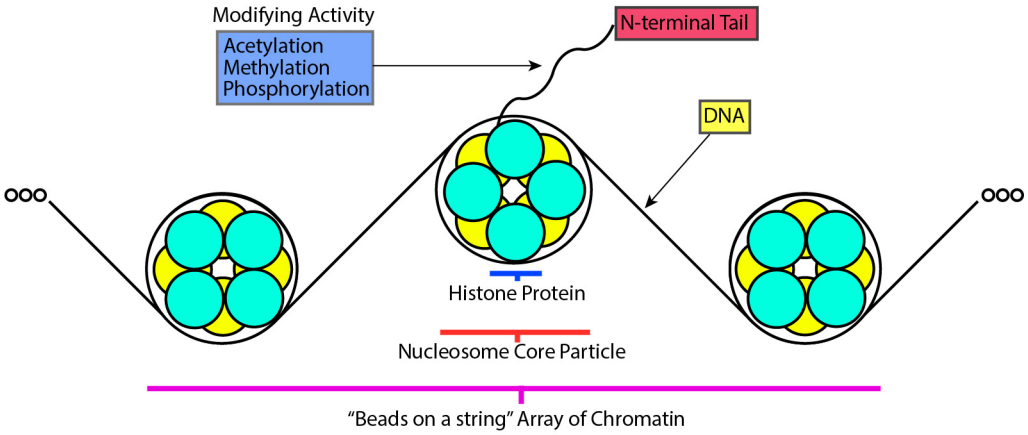

Figure 1.2: A generic schematic of the "beads on a string" arrangement of chromatin, typically found in eukaryotic systems. Histone octamers form a platform for DNA to wrap around. The nucleosome core particle is the basic repeating unit of chromatin, comprised of 147 bp of DNA wrapped around a histone octamer. Post-translational modification of the N-terminal tails of these histones, which protrude from the NCP, is completed by large, multisubunit protein complexes. These modifications can alter the packaging of chromatin (heterochromatin vs euchromatin), or act as molecular markers that allow for the association of more modifying complexes to specific loci.

1.1.1 Transcription

In eukaryotes, RNA Polymerase II (or Pol II) is responsible for transcribing gene loci into mRNA, which is then exported from the nucleus and translated by the ribosome into protein (the process of splicing; removing introns, or non-coding RNA, from the mRNA sequence must also occur prior to translation) (Lee and Young 2000).

Pol II is a large, multisubunit protein complex that has been thoroughly studied in terms of its regulation and structure (Bernecky et al. 2016, Ehara et al. 2017). The complex initiates transcription at specific sites, upstream of coding DNA, termed promoters. Recruitment of Pol II to promoters is facilitated by general transcription factors (and other factors for recruitment to specific loci) to form a pre-initiation complex. The complex consists of; TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH (which contains the TATA box binding protein, TBP) and the mediator complex (Kelleher et al. 1990, Louder et al. 2016, Harper and Taatjes, 2018.)(**Figure 1.3**).

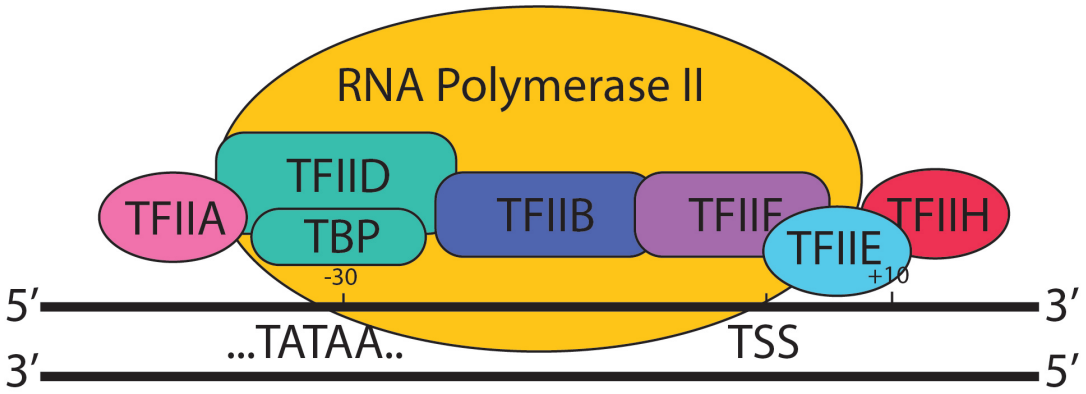

Figure 1.3: Schematic of the pre-initiation complex over a eukaryotic promoter sequence. RNA Pol II forms a large complex with general transcriptions factors TFIIA-H. The TFIID complex associates with a TATA box sequence, approximately 30 bp upstream of the transcription start site (TSS) via the TBP subunit. Other subunits help recruit the polymerase and prime the promoter DNA for transcription initiation.

Pol II is recruited to specific gene loci via the mediator complex, a multisubunit complex that bridges the PIC to transcription factors, that associate with regions upstream of the promoter (enhancers). This provides an additional layer of spatial and temporal regulation of gene activation and is further controlled via the modification of histones, that promotes/represses the assembly of the PIC.

Upon initiation, the polymerase escapes the initiation complex, and the following processes of transcription elongation and termination, are tightly regulated. Much of this regulation is implemented by the phosphorylation of the Pol II C-terminal domain (CTD), as well as the association of various factors at specific intervals.

1.1.2 Readers, Writers and Erasers

Advances in the fields of structural, molecular and cell biology, have revealed that post-translational modifications (PTM) of histones, play a significant role in chromatin structure and by extension, the regulation of cellular processes such as transcription. Post-translational modifications are carried out by coactivator complexes, which are also tightly regulated both spatially and temporally, ensuring the targeted activation/inhibition of target gene loci (Kouzarides, 2007).

The majority of PTMs occur on the N-terminal tails of histones, which are often protruding out from the NCP and are therefore the most accessible region of the protein assembly (**Figure 1.2**).

There are several types of PTM, including; phosphorylation, methylation and acetylation, to name a few (Sterner & Berger 2000, Zhang & Reinberg 2001, Zhang et al. 2014). Many of these modifications are reversible and a set of antagonistic complexes are in place, to remove said modifications at the appropriate time and position.

Post-translational modification of histones can directly influence the structure of chromatin, or act as markers for binding and additional interactions with other chromatin modifying complexes. This allows for dynamic regulation of gene expression, in what has been termed: the histone code (Jenuwein et al. 2001).

Coactivator complexes can also be recruited to upstream activator sequences/enhancer regions by transcription factors and promote/repress the assembly of basal transcriptional machinery (**Figure 1.4**).

Chromatin must also be modified after transcription/replication/repair events have been initiated. For example, remodelling complexes such as FACT modify the histone composition of nucleosomes, thereby destabilising the NCP in preparation for incoming Pol II, maintaining transcription elongation rates (Belotserkovskaya et al. 2003). The FACT complex removes the H2A-H2B dimer, acting as a histone chaperone.

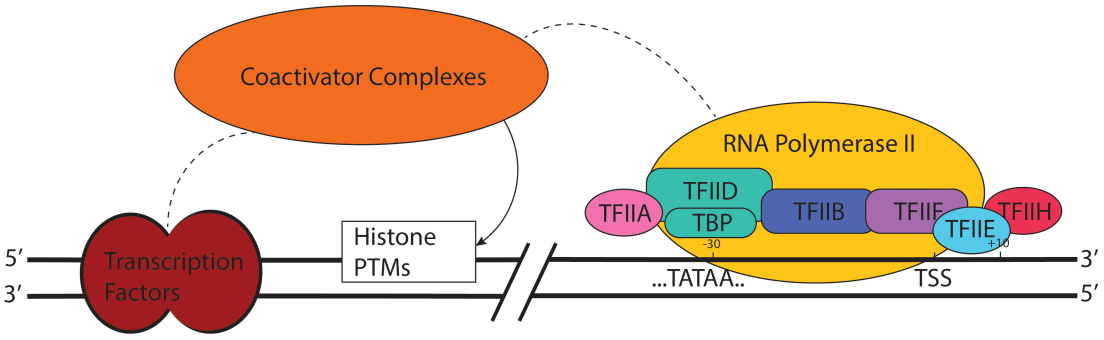

Figure 1.4: Schematic summary of coactivator recruitment. Transcription activators bind upstream of transcription initiation sites. Coactivators are subsequently recruited and carry out post-translational modification of histones and/or directly interact with the forming pre-initiation complex, promoting transcription activation.

1.1.3 The Role of Acetylation

Histone acetylation results in the incorporation of an acetyl functional group onto lysine residues, from an acetyl-CoA substrate, resulting in loss of the overall positive charge of the NH_3^+ group.

Acetylation occurs at lysine residues located on the N-terminal tail of histones and was initially proposed to reduce the overall positive charge on the surface of nucleosomes, resulting in reduced electrostatic interactions with the phosphate backbone of DNA. The net result is the loosening chromatin and exposing promoters to transcriptional machinery, for increased levels of expression (Loidl 1994).

Advances in our understanding of the histone code, have shown that these acetylation markers, also act as recruiting interfaces for transcription factors and DNA repair machinery (Natsume-Kitatani et al. 2011, Tscherner et al. 2012). It is likely that both changes to the physico-chemical properties (as suggested by Loidl et al.) in combination with the recruitment of transcription factors, drive increased levels of transcription.

Histone acetylation occurs at specific amino acid residues on N-terminal tails, which protrude out from the complex and are more accessible to coactivator complexes. Histone acetyltransferase complexes, such as SAGA and NuA4, are responsible for this form of modification. These acetylation events have a wide range of biological consequences, some of which are catastrophic in the event of misregulation.

1.2 NuA4

The first indication that histones were acetylated in a targeted fashion was established by Kuo et al. (1996), who showed that the histone acetyltransferase (HAT) Gcn5, acetylated lysine residues on the N-terminal tails of H3 and H4.

In search of more HAT complexes, Grant et al. (1997) began to isolate fractions produced from yeast cell extract, which contained HAT activity (these fractions were produced using a combination of ion exchange and gel filtration). HAT activity was measured by incubating fractions with purified nucleosome (or free histones) and tritium labelled acetyl-CoA. After incubation, samples were separated by SDS-PAGE and acetylated histones could be visualised by fluorography and identified by their molecular mass on the polyacrylamide gel.

The final size-exclusion chromatography step, produced four distinct peaks representing protein complexes with HAT activity. Only two of these peaks contained the Gcn5 enzyme, though Grant et al. primarily focussed on the Gcn5 containing protein complex (called SAGA).

Allard et al. (1999) would make the first attempt to characterise these newly discovered HAT complexes, noticing that one could tetra-acetylate H4 and H2A, a profile that matched the already characterised HAT enzymes Tip60 and yeast homolog Esa1 (Yamamoto & Horikoshi 1997). By raising antibodies against Esa1 and incubating their purified HAT fraction with a resin cross-linked to the antibody, they were able to capture the complex and confirm that Esa1 was a component of this novel multi-protein fraction.
Tip60 and Esa1 are members of the MYST family (MOZ, Ybf2/ Sas3, Sas2 and Tip60) of HATs. This is the largest of the HAT families and has been implicated in cancer pathology, cell cycle progression and even chromosomal translocation, when investigating certain forms of Leukemia (Avvakumov & Côté 2007).

Esa1 is the only HAT enzyme found to be essential in yeast, though mutations that abolish *in vitro* activity are not lethal (Smith et al. 1998). It was suggested that additional interacting proteins aided in stabilising the enzyme, as with Gcn5 and the other SAGA subunits (e.g. Ada2), which could compensate for the reduced activity of the catalytic subunit *in vivo*. Temperature sensitive(*ts*) mutants of Esa1 results in loss of H4 and H2A acetylation and prevented the purification of the novel "NuA4" complex (nucleosome acetyltransferase of H4).

NuA4 and SAGA acetyltransferase complexes preferentially associate with promoters, this association is followed by increased levels of acetylation in these regions and transcriptional activation of downstream sequences (Utley et al. 1998, Ginsburg et al. 2009). The association to these regions by either complex, is driven by the interaction between the shared subunit; Tra1 and transcription factors, such as HAP4 (Brown et al. 2001).

The last two decades of research has revealed that the *S. cereviseae* NuA4 is a thirteen subunit complex, with multiple subcomplexes of varying function. This includes: global levels of acetylation, recruitment to transcription factors, promotion of transcription elongation rates, splicing and DNA repair. Figure 1.5 shows a schematic, summarising our current understanding of NuA4 subunit composition, with subcomplexes colour coded according to their *in vivo* role.

At present, there are no high-resolution structures of the complete complex. Chittuluru et al. (2011) were able to provide a low-resolution structure of the complete complex using electron microscopy, to a final resolution of 20 Å. This has since been shown to actually correspond to a low-resolution structure of the single Tra1 subunit (Diaz-Santin et al, 2017), as evidenced by recent incomplete EM structures of the NuA4 complex, that will be discussed at the end of this chapter.

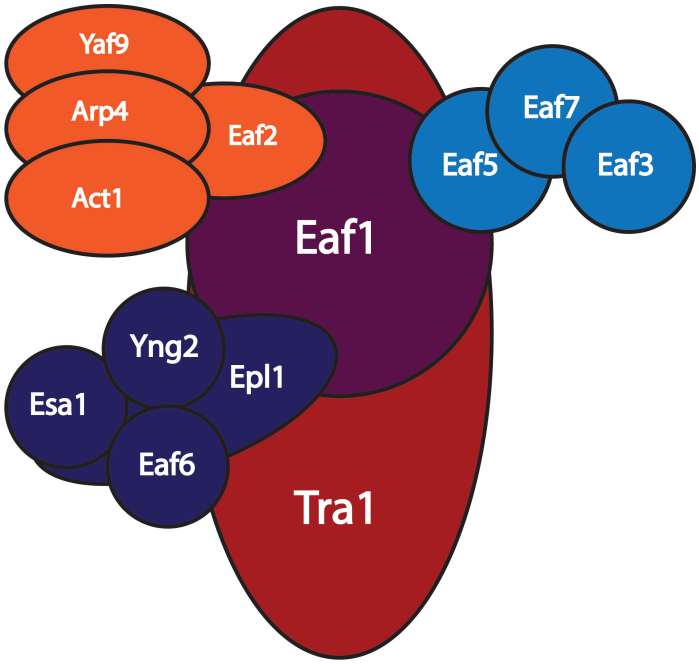

Figure 1.5: A schematic representation of the yeast NuA4 complex. Subunits are arranged according to their interactions and colour-coded according to their proposed biological function. Epl1, Yng2, Eaf6 and Esa1 are coded dark blue and represent the Piccolo NuA4 subcomplex, which houses the catalytic Esa1. The central scaffold protein Eaf1 is coded purple and is the only NuA4 specific protein found in yeast. Tra1 is the largest component of the 1.3 MDa complex, constituting approximately 40% of the molecular mass of NuA4 and shown to physically link the complex to transcriptional activators such as c-Myc. Eaf3/5/7 are coloured light blue and represent the TINTIN (trimer independent of NuA4 for transcription interactions with nucleosomes) subcomplex. Subunits that are shared with the SWR1 chromatin remodelling complex and implicated in DNA damage repair mechanisms, have been labelled orange.

1.2.1 The Tip60 Complex

Doyon et al. (2004) is a good summary of all that was known about NuA4 at the time. They note that NuA4 shares 12 of its 13 subunits with the human Tip60 HAT complex (with the exception of Eaf5, which is unique to *S. cerevisiae*) (**Figure 1.6**). The human Tip60 complex is a fusion of the yeast NuA4 acetyltransferase and the SWR1 remodelling complex (which also shares four subunits with NuA4) and possesses the functionality of both yeast complexes.

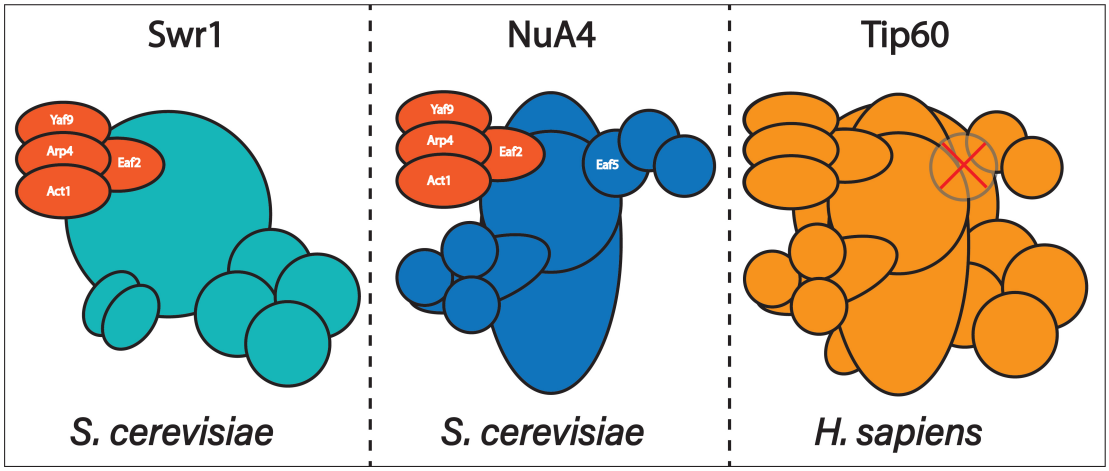

Figure 1.6: A schematic representation of yeast NuA4, SWR1 and human Tip60 complexes. Yeast NuA4 shares 12 of its 13 subunits with the human Tip60, which does not possess an Eaf5 ortholog. NuA4 and SWR1 are known to possess a shared subcomplex, which has been highlighted in orange. The Tip60 complex appears to be a fusion of the yeast HAT NuA4 and the SWR1 remodelling complex and has been reported to show both activities *in vivo*. Tip60 functions as an acetyltransferase and remodelling complex. The Tip60 complex has been implicated in Alzheimer's and cancer pathologies.

Figure 1.6 highlights the benefits of using yeast NuA4 as a proxy for understanding the Tip60 complex function. While Tip60 shows additional chromatin remodelling activity (due to SWR1 subunit orthologs), later reports would show that yeast NuA4 and SWR1 work in concert, to acetylate and then remodel chromatin, similar to the activity of the Tip60 complex (Zhang et al. 2004).

This coordination is achieved in yeast, in part, by the two complexes sharing subunits. The role of the Yaf9 subunit is a good example of the two complexes working together, as it enables the incorporation of non-canonical histone H2A.Z in the promoter region of the *PHO5* gene. This alteration is usually associated with regions of chromatin being prepared for transcriptional activation and is often enriched near hyperacetylated H4 (likely a reflection of SWR1 and NuA4 joint modification of chromatin) (Santisteban et al. 2000, Kurdistani & Grunstein. 2003).

Tip60 has been shown to be essential for DNA damage response pathways, which is likely due to its role in acetylation (via the NuA4-like subunits) and remodelling (Ikura et al. 2000). These activities possibly prepare regions of the genome in close proximity to sites of DNA damage, for repair machinery binding, in a similar manner to how chromatin is primed for the formation of the pre-initiation complex, for transcriptional activation. However, the exact function of Tip60/NuA4 in DNA damage repair mechanisms remains largely uncharacterised, while its role as a coactivator complex has slowly been delineated.

1.2.2 The Piccolo Subcomplex

In 2003, Boudreault et al. showed that Esa1, Epl1 and Yng2 could form a stable trimer *in vivo* which was catalytically active and likely responsible for global, non-targeted acetylation of H4. Transient inhibition of the expression of any of these proteins (when under the influence of a GAL1 promoter), led to the accumulation of cells in G2/M stage of the cell cycle (measured by fluorescence-activated cell sorting).

This subcomplex was named Piccolo and all three components are essential for cell cycle progression. The transient repression experiments suggested that global H4 acetylation is carried out by the Piccolo subcomplex, as opposed to the targeted acetylation carried out by complete NuA4. This global activity is independent of the rest of the NuA4 complex.

Mutations in Piccolo subunits result in sensitivity to rapamycin and MMS, implying the complex plays a role in ribosome biogenesis, TOR signalling and DNA damage repair.

Piccolo acetylates residues K5, K8, K12 and K16 of H4 with approximately 20X higher rate in the complete NCP, compared to the free histone (Berndsen et al. 2007). Comparing free Esa1 acetylation activity to its activity in the Piccolo trimer, reveals that the complex acetylates nucleosomes preferentially, whereas the free enzyme is incapable of acetylating the NCP and will acetylate H4 with equal activity and efficiency, as it does peptides mimicking the N-terminal tail of H4. This data supports the roles of Yng2 and Epl1 in both facilitating NCP binding and increased catalytic activity from Esa1.

Yan et al. (2000) would provide the first crystal structure of the active site of Esa1 in complex with acetyl-CoA. They identified the region critical for HAT activity *in vitro* by generating a series of truncated Esa1 constructs, narrowing their search down to residues 160 - 435.

While these constructs retained catalytic activity *in vitro*, their *in vivo* activity appeared to be disrupted, resulting in growth defects during complementation studies. They speculated that the deleted regions contained interfaces for interactions with other components of NuA4, which were required for fully functioning activity.

Superimposing Esa1 with the structure of Gcn5 (PCAF) and Hat1 bound to substrate, revealed that acetyl-CoA binds to a similarly structured position in each case and was in close proximity to a conserved glutamate residue, a proton acceptor during the catalytic mechanism. They also noted that there was sufficient space for the side chain of a lysine residue being modified, to dock between the glutamate and the acetyl-CoA binding positions.

While useful, this structure was of the incomplete enzyme and NuA4 at this point was still a relatively obscure complex in terms of our understanding of subunit composition; with only Tra1 and Esa1 having been identified. More information was required to deduce the apparent specific transcriptional activation associated with NuA4's *in vivo* activity.

Esa1 contains an essential non-canonical chromodomain (located between its N-terminal and catalytic domain) which is unable to bind methylated lysines (Selleck et al. 2005, Jacobs et al. 2001). The chromodomain identified would later be shown to actually form an anti-parallel 5-stranded β-barrel structure, more akin to a tudor domain than canonical chromodomains (Shimojo et al. 2008, Maurer-Stroh et al. 2003).

The tudor domain shows structural similarity to the archael DNA binding protein, SSo7D and was shown to promote nucleosome binding, potentially via direct contact with DNA (Huang & Tan, 2013) (**Figure 1.7**).

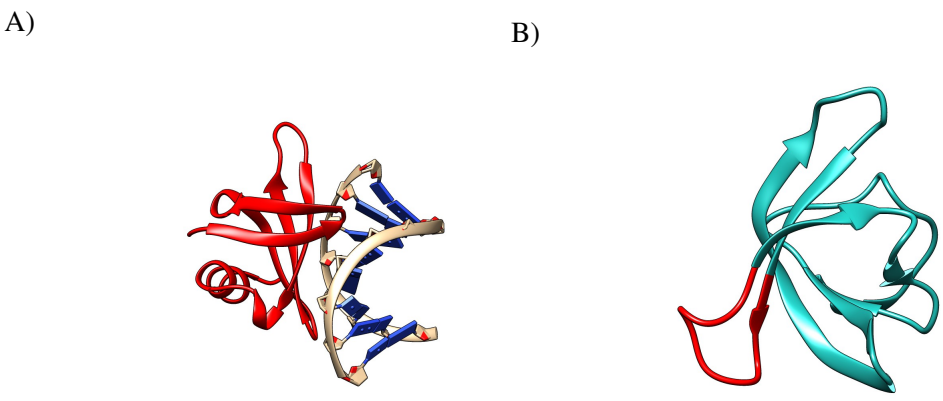

Figure 1.7: A) 2.1 Å structure of Archaeal Sso7D protein (red) in complex with double stranded DNA (PDB entry: 1BNZ). **B)** NMR solution structure of of the Esa1 tudor domain (cyan). The domain forms an anti-parallel β-barrel with a distinct loop (red) conserved between Esa1 and MOF, which has been implicated in the catalytic activity of the enzyme, despite being distinct from the active site. (PDB entry: 2RO0).

Yng2, a plant homeodomain (PHD) containing protein, is a component of NuA4 (Nourani et al. 2001, Eisen et al. 2001). The PHD domain is a cys-his-cys motif; found in many proteins involved in chromatin modification and implicated in cancer pathology (Aasland et al. 1995, Gunduz et al. 2000). Yng2 mutants are also growth sensitive to MMS, a DNA damaging agent (Garkavtsev et al. 1996). This is to be expected, when we consider it shares sequence similarity with the tumour suppressor ING1.

Interestingly, this PHD domain is not essential for restoring the wild-type (WT) levels of sensitivity to DNA damaging agents, which leaves questions regarding the domain's overall role. Like its ortholog ING1, Yng2 directly interacts with human "guardian of the genome" p53 and this interaction results in the hyperacetylation of nucleosomes in close proximity to p53.

The Epl1 subunit interacts directly with the other two components of Esa1 and Yng2, but these two proteins do not co-purify in the absence of Epl1, strongly supporting its role as the scaffold subunit within the trimer. Strains expressing Epl1 truncated at the non-essential C-terminal domain, prevent co-purification of NuA4 via Esa1, but cells are otherwise viable, suggesting that this smaller subcomplex is the essential component of NuA4 for cell viability.

A 20-residue N-terminal region, in the conserved EPcA (Enhancer of poylcomb A) domain of Epl1, is essential for HAT activity (Boudreault et al. 2003). Like the Esa1 tudor domain, mutations in this region reduce catalytic activity, as opposed to the binding of nucleosomes (Huang & Tan. 2013).

Rossetto et al. (2014) were able to to show that Eaf1 deletion followed by TAP-tagged purification via Eaf6, yields the three subunits of Piccolo. C-terminal truncation of Epl1, which results in separation between Piccolo and the rest of NuA4, followed by Eaf6 TAP-tag purification yields identical results, indicating that Eaf6 is a member of the Piccolo subcomplex.

This is a reasonable suggestion, when we consider that the human ortholog of Eaf6 forms a tetrameric complex *in vivo*, which is responsible for global acetylation of nucleosomes in humans (though not with the Tip60 HAT) (Doyon et al. 2006). Piccolo has since been structurally characterised to a high resolution, in the form of individual structures of Esa1 (in complex with its acetyl-CoA substrate) and the complete complex (Yuan et al. 2012, Xu et al. 2016).

The major leap in Piccolo structure/function understanding was recently made by Xu et al. (2016), who produced high-resolution structures of Piccolo (including Eaf6) in complex with acetyl-CoA and a peptide mimicking the N-terminal tail of Ht1z. The peptide bound structure was produced at a resolution of 3.23Å(R_{free}/R_{work} = 25.0/27.3) and includes all essential domains. (**Figure 1.8**)

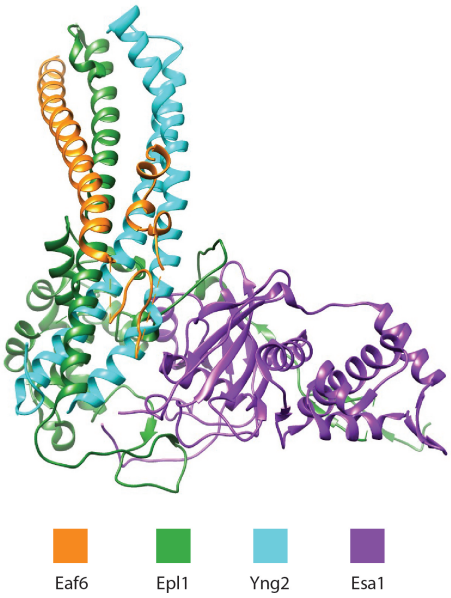

Figure 1.8: Subunits of the Piccolo complex, from a 3.23 Å resolution crystal structure (R_{free}/R_{work} = 25.0/27.3). PDB entry: 5J9Q. Includes most critical domains for integrity and HAT activity, with the exception of tudor domain of Esa1. Subunits have been colour-coded and are arranged broadly into two distinct regions; a catalytic core consisting of Esa1 and the EPcA-I domain of Epl1, and a helical bundle of Eaf6, Yng2 and the EPcA-II domain of Epl1.

The structure can be broadly split into two regions; the catalytic core and a four-helix bundle. The catalytic core consists of Esa1 and an N-terminal end of the EPcA domain of Epl1, while the C-terminal region of the EPcA domain forms a four-helix bundle with the remaining Piccolo subunits and likely fulfils a structural stabilising role in the complex.

The two distinct regions within the EPcA domain show weak interactions and are connected by a loop described as the dual-function loop (DFL). The N-terminal region of the EPcA domain, which had been characterised by Huang & Tan (2013) for its role in catalytic activity, extends out from the complex to then loop around Esa1 and an Epl1 globular domain, that interacts with Esa1, at a site neighbouring the catalytic domain.

The cleft between a conserved glutamate and CoA binding site, forms the docking position for histones (specifically the amine side chain of lysine) and resides within a α2-β7 loop, termed the histone binding loop (HBL).

They show that, unlike Yan et al.'s structure of free Esa1, the α2-helix of Esa1 in Piccolo interacts with the DFL via a FRR motif, which rearranges the loop to extend outwards. Additional points of contact with Epl1 and Esa1 ensures that the Esa1 HBL loop stays in this extended conformation. They confirm an auto-acetylation event as proposed by Yuan et al. (2012), which is further stabilised by interaction with Ser291 and Tyr289.

The lysine binding pocket is shallow and substitution of amino acids downstream from the lysine on the peptide, to larger residues (such as histidine) produces a dramatic reduction in catalytic activity. H4 N-terminal lysines are usually flanked by relatively short residues such as alanine, which show minimal steric clashing with the residues of the Esa1 binding pocket, this is the basis for preferential H4 acetylation by NuA4.

Finally, Xu et al. produced a 7.9 Å resolution structure of Piccolo in complex with the NCP, by docking their newly produced structure, the tudor domain solution structure produced by Shimojo et al. and a high-resolution NCP structure onto their electron density.

The helical bundle of the Piccolo structure, appears to protrude at a peripheral position to the NCP and the conserved glutamate of Esa1, rests 24 Å from the N-terminal tail of H4, which is sufficient for access to the four lysine residues. The H2A N-terminal tail on the other hand, is nearly 3X the distance from this glutamate as the H4 tail is, likely exacerbating the preferential acetylation of H4.

H2B and H3 N-terminal tails by contrast, are on another face of the nucleosome structure, preventing direct interaction with the active site. This is direct evidence of what had been inferred by many years of biochemistry research; that the structures of the Piccolo complex and NCP, positions Esa1 in such a way that is optimal for H4 acetylation, as opposed to other histone proteins.

1.2.3 Tra1 and NuA4 Recruitment

SDS-PAGE analysis of the NuA4 fractions consistently produce a high-molecular weight band (400 kDa). Tra1 was quickly identified as a candidate protein; being one of the few yeast proteins to match this MW and is also a component of the HAT complex, SAGA (Saleh et al. 1998).

Tra1, like its human ortholog TRRAP, had been shown by Saleh et. al to associate with transcription factors such as c-Myc, suggesting a direct connection between chromatin acetylation and transcription activation.

Other groups had shown that Tra1 was essential for c-Myc and E1A driven transcription activation, as Tra1 mutants blocked oncogenic transformation of the transcription factors (McMahon et al. 1998).

Tra1 likely plays additional roles in the mechanism of SAGA and NuA4 activity, when we consider mutations that do not alter the targeting or integrity of either coactivator complex, but nonetheless show reduced rates of transcription (Knutson and Hahn, 2011).

The subunit is part of the Phosphoinositide 3-Kinase-related kinase (PIKK) family of proteins, which play a role in growth regulation, mTOR signalling and cancer pathology when misregulated (Yang et al. 2013, Baretic and Williams, 2014).

Tra1(and human TRRAP) lacks catalytic activity and its structure has been solved to 3.7 Å resolution by cryo-EM (Diaz-Santin et al. 2017)(**Figure 1.9**).

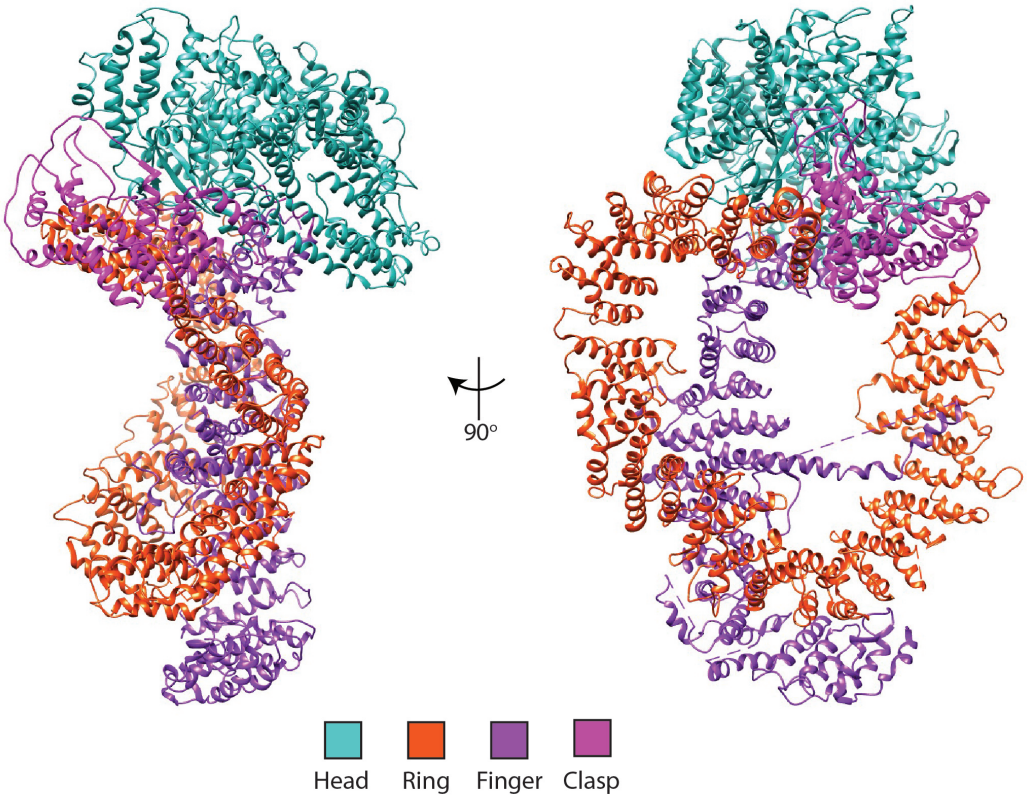

Figure 1.9: 3.7 Åresolution cryo-EM structure, of the PIKK family protein, Tra1 (PDB = 5OJS). The protein is structured as a "diamond ring", with the inactive kinase, FRB and FAT domains forming the head of the structure. The majority of the structure however is alpha-helical, arranged into HEAT repeat solenoids, found in the "finger" and "ring" regions.

The protein contains all domains associated with the PIKK family (Baretic and Williams 2014, Aylett et al. 2016), including an FRB domain, the inactive catalytic site and a FATC domain at the C-terminal. Over 80% of the structure is comprised of α-helical HEAT repeat domains, these solenoid structures form a large "ring" and "finger" region, which rests beneath the "head" and "clasp" regions. HEAT repeats are found throughout the structure, but the head and clasp domains contain the FAT, FRB and inactive kinase domains.

Comparisons to other PIKK family members reveals that Tra1 shows mutations in ATP/Mg binding sites and contains an 18 residue insertion in a known activation loop, which may explain its inactivity (Yang et al. 2013, Sibanda et al. 2017).

The ring and finger regions contains high-surface area, solvent accessible channels, though activator binding sites which were previously identified by mutation studies, were found scattered across the entire complex (Brown et al. 2001, Knutson and Hahn, 2011). The wide distribution of these sites ensures that mutations that would disrupt binding between Tra1 and one activator, will not interfere with binding to other activators.

1.2.4 Eaf1: The NuA4 Scaffold

Eaf1 is the essential NuA4 scaffold, around which other subunits in the complex assemble (Auger et al. 2008)(**Figure 1.5**). Eaf1 shows strong identity with the Tip60 component p400/Domino, though lacks a SWI/SNF family ATPase domain, essential for the incorporation of the non-canonical histone H2A.Z into the NCP (Cöté et al. 1994, Narlikar et al. 2013).

TAP purification of natively expressed Eaf1, followed by mass spectrometry and western blot analysis, identified Eaf1 as the only subunit specific to the complete NuA4 complex.

TS mutants show sensitivity to MMS and ionising radiation, further implicating NuA4's role in DNA repair mechanisms. Like the subunits of Piccolo-NuA4, Eaf1 mutants also show rapamycin sensitivity and reduced *PHO5* expression levels (Bennett et al. 2001, Auger et al. 2008).

Eaf1 makes direct contact with Epl1, tethering the Piccolo subcomplex, as well as Tra1 via a C-terminal SANT domain. SANT domains are usually found in proteins that play a role in chromatin modification and often possess histone or DNA binding properties (Boyer et al. 2004). Eaf1 also makes contact with NuA4 subunits shared with the SWR1 remodelling complex (Eaf2, Arp4, Act1 and Yaf9) via Eaf2 and Rossetto et al. (2014) would show that the Eaf5/7/3 subcomplex would also tether to Eaf1 via Eaf5. These contacts were inferred by several studies and finally verified with by the cryo-EM structure of the core NuA4 complex Wang et al. 2018) (**Figure 1.10**).

While a functioning Piccolo subcomplex is the minimal requirement for viable HAT activity *in vivo*, Eaf1 couples this non-specific HAT activity to other NuA4 components that act to guide the coactivator complex to specific loci (Tra1 and Arp4 for example). The loss of Eaf1, while not immediately catastrophic for the cell, is required for the specific activation of transcription for certain genes. The subsequent acetylation of promoters that follows, results in recruitment of basal transcription machinery, such as TFIID (Uprety, Sen and Bhaumik, 2015).

1.2.5 Chromatin Remodelling Module

Eaf2, Arp4, Act1 and Yaf9 are all components of the SWR1 chromatin remodelling complex and NuA4, (Galarneau et al. 2000, Le Masson et al. 2003, Auger et al. 2008). The role of these subunits in NuA4 is the least understood, though some efforts have been made to delineate their function.

Eaf2 tethers this module to Eaf1 and makes direct contact with Yaf9. As with Eaf1, Eaf2 contains a SANT domain and is essential for WT levels of cell growth.

Mutations in Eaf2 also result in MMS sensitivity, but unlike other NuA4 components, these do not show sensitivity to rapamycin.

Yaf9 mutants produce growth defects and reduced H4 acetylation levels, and is also implicated in spindle pole body (SPB) formation (Loewith et al. 2001, Gavin et al. 2002, Le Masson et al. 2003). While mutations in Yaf9 show sensitivity to benomyl (a microtubule disrupting compound), it is difficult to determine whether these mutations (and their consequences) are a result of disruption to NuA4 activity or SWR1.

Act1 had already been shown to be a component of the mammalian chromatin modifying complex SWI/SNF; responsible for altering the histone composition of canonical NCP (Zhao et al. 1998).

Arp4 is an interesting component, having been shown to play a role in activation of specific loci, such as *LYS2* (Jiang & Stillman 1996) and directly interacts with histones (Galarneau et al. 2000). *TS* mutations of Arp4 have similar results to mutations in Yaf9 and Esa1, leading to reduced growth and even reduced levels of acetylation in the same gene loci.

The above mentioned data implicates this module in DNA repair functionality, though it is difficult to distinguish this from consequences for the SWR1 complex, which is also important for DNA repair processes (Rosa et al. 2013).

1.2.6 Core NuA4 and SAGA

Throughout the course of this project, advances have been made in our understanding of the structure and function of the complex (discussed in the next section), which has been aided largely due to advances in electron microscopy techniques and the so called "resolution revolution."

Figure 1.10 shows a 4.7 Å resolution cryo-EM structure of the NuA4 TEEAA-complex (Tra1, Eaf1, Eaf5, Actin, and Arp4) by Wang et al. (2018).

This represents the most complete (and highest resolution) image of the NuA4 complex available. While they are able to purify the complete 13 subunit complex, the predominant complex observed in their microscopy experiments was deficient of Eaf7, Eaf3, Swc4(Eaf2), Yaf9 and Piccolo (though they also produce a 7.8 Å structure of this "core" complex, with a docked Piccolo crystal structure).

Their structures suggest that Tra1 also plays a scaffolding role in parallel with Eaf1, making contacts with Eaf5, as well as the two Piccolo subunits; Epl1 and Yng2.

Wang et al. observe direct contact between the Eaf1 HSA domain and the actin, and Arp4 subunits (as had previously been inferred by Auger et al. 2008) via a series of hydrophobic interactions. The Eaf1 SANT domain was also shown to be in direct contact with Tra1's LBE and FACT domains, by hydrogen bonding and ionic interactions, partially explaining Tra1 mutations associated with reduced growth rates in yeast (Hoke et al. 2010). Wang et al. (2018) produced a lower resolution structure of TEEAA-Piccolo, suggesting that the catalytic subcomplex makes contact with Tra1 via Yng2 and with the Eaf1 HSA domains, via the C-terminal of Epl1.

Eaf1 in the TEEAA structure has relatively large regions that are unresolved/disordered, but upon binding of the Piccolo complex, the scaffolding complex becomes more compact/ordered.

This TEEAA subcomplex gives some definitive insights into subunit arrangement, and is generally in agreement with what had been inferred indirectly in the previously discussed literature. The structure of the Tra1 subunit is largely in agreement with the structure our group had previously produced (**Figure 1.10**).

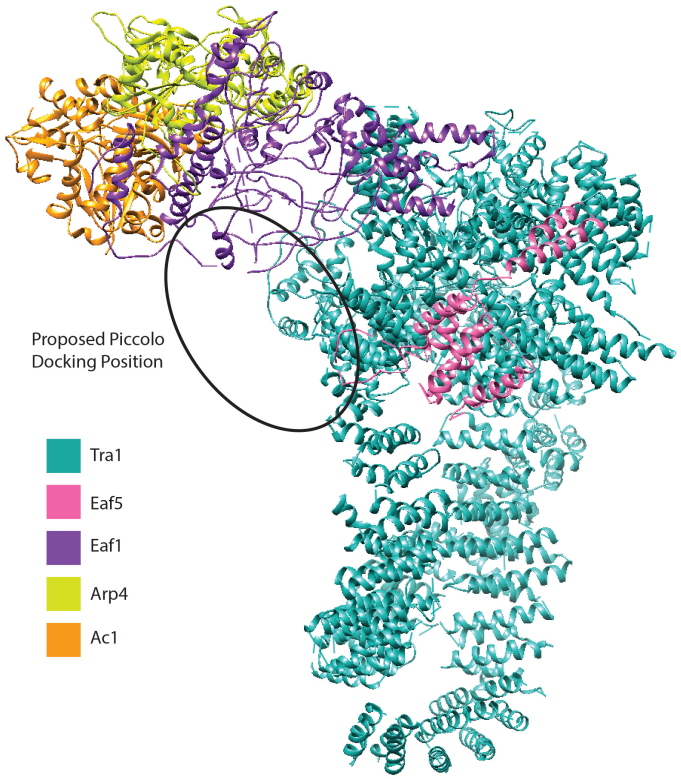

Figure 1.10: 4.7 Å resolution cryo-EM structure of the TEEAA (Tra1, Eaf1, Eaf5, Act1, Arp4) NuA4 subcomplex produced by Wang et al. (2018). An additional structure of lower resolution was also produced (data not included), docking the position of the Piccolo subcomplex, its docking position in the TEEAA complex has been circled. Significant portions of the Eaf1 subunit remain disordered/unresolved, however both SANT and HSA domains, which are responsible for direct contacts with Tra1 and Arp4/actin, respectively, are represented.

A complete, high-resolution picture of the NuA4 complex remains elusive, in part due to the flexibility of key modules, and also due to the transient nature of interactions between specific interfaces. Moreover, recent data has emerged that conflicts with the structure of the TEEAA (Setiaputra et al. 2018, Cheung and Diaz-Santin, 2019)(**Figure 1.11**). Negative stain EM and cross-linking mass spectrometry data from Setiaputra et al. (2018), are in general agreement with the structure of the TEAA complex, though they suggest a different position for the Piccolo subcomplex within the greater NuA4 complex.

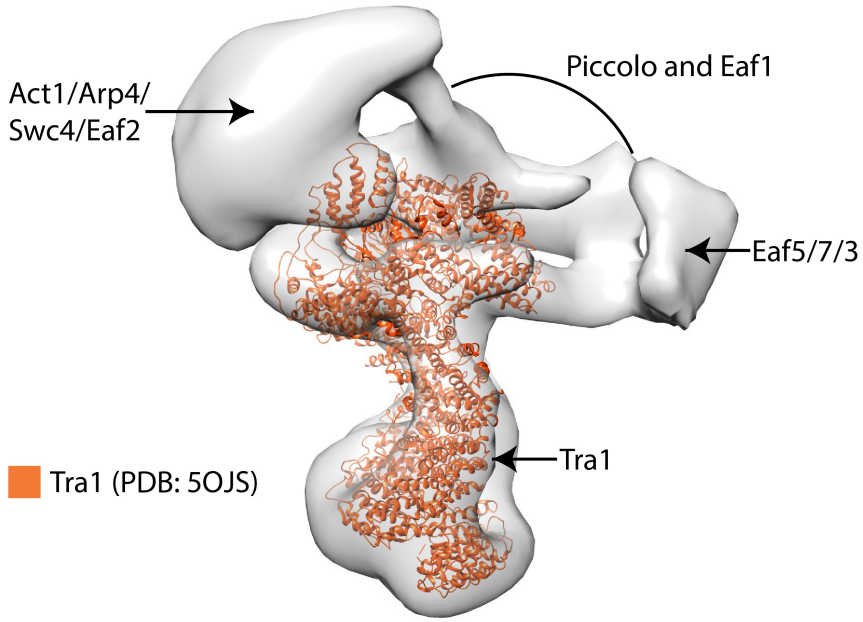

Figure 1.11: Negative stain data, proposed to depict the structure of NuA4 (Accession: EMD-7131) with a higher resolution, cryo-EM structure of Tra1 (PDB = 5OJS) superimposed. The centre of the complex is believed to house Tra1, Eaf1 and the Piccolo subcomplex. Two lobes, one suspected to correspond to Eaf5/7/3 and the other corresponding to Ac1/Arp4/Swc4/Eaf2, can be seen extending away from the core of Eaf1 and the head region of Tra1.

While NuA4 is slowly resolved, other coactivators, such as SAGA are more clearly defined (Liu et al. 2019). Also a HAT complex, implicated in global transcription activation and recruited to promoters via Tra1, SAGA is a larger complex (1.8 MDa) with additional deubiquitylation activity.

However, recent studies suggest that the SAGA complex should be considered a general transcription factor in *S. cerevisiae*, after it was found to bind to upstream activator sequences (UAS) of almost all Pol II transcribed genes, occupying the same region as the essential Mediator complex (Baptista et al. 2017). As our understanding of NuA4's function and recruitment improves, it would be interesting to compare to SAGA, considering the complexes share such a large and versatile recruitment interface in Tra1.

1.3 The TINTIN Subcomplex

1.3.1 Discovery and Basic Properties

TINTIN (**t**rimer **i**ndependent of **N**uA4 for **t**ranscription **i**nteractions with **n**ucleosomes) is the most recently discovered subcomplex of NuA4, comprising of Eaf3, Eaf5 and Eaf7 (Rossetto et al. 2014, Bhat et al. 2015). The complex can be natively purified upon affinity tagging any of the three subunits and when followed by gel filtration, produces fractions of complete NuA4 and a distinct peak, containing the three subunits.

1.3.2 Composition and Structure

Eaf5 is unique to yeast and is believed to serve as a bridge between Eaf3/Eaf7 and NuA4, though has no recognised conserved domain within its sequence (Rossetto et al. 2014). Eaf5's structure has been partially characterised when in complex with the TEEAA "core" NuA4 complex, via extensive contacts with Tra1 (**Figure 1.12**). Much of Eaf5 remains unresolved, including its points of contact with Eaf1, that was identified by cross-linking MS. An interaction interface with the two other TINTIN subunit has also yet to be identified (Xie et al, 2015).

Eaf3 is a chromodomain containing protein, that is shared with the histone deacetylase (HDAC) Rpd3S, which is antagonistic to NuA4. Deletion of the subunit results in no growth phenotype and the NuA4 complex appears otherwise stable, but reduced levels of transcription are reported, originally proposed to be a consequence of a shift in acetylation patterns (Eisen et al. 2001, Reid et al. 2004).

The chromodomain preferentially associates with trimethylated lysine 36 (K36) of H3, a post-translational modification carried out by the Set2 complex, during transcription elongation (Sun et al. 2008).

The domain has been structurally characterised, revealing a distinct "chromo barrel" structure, similar to the human ortholog of Eaf3, MRG15 (**Figure 1.13**)(Zhang et al. 2006 , Sun et al. 2008).

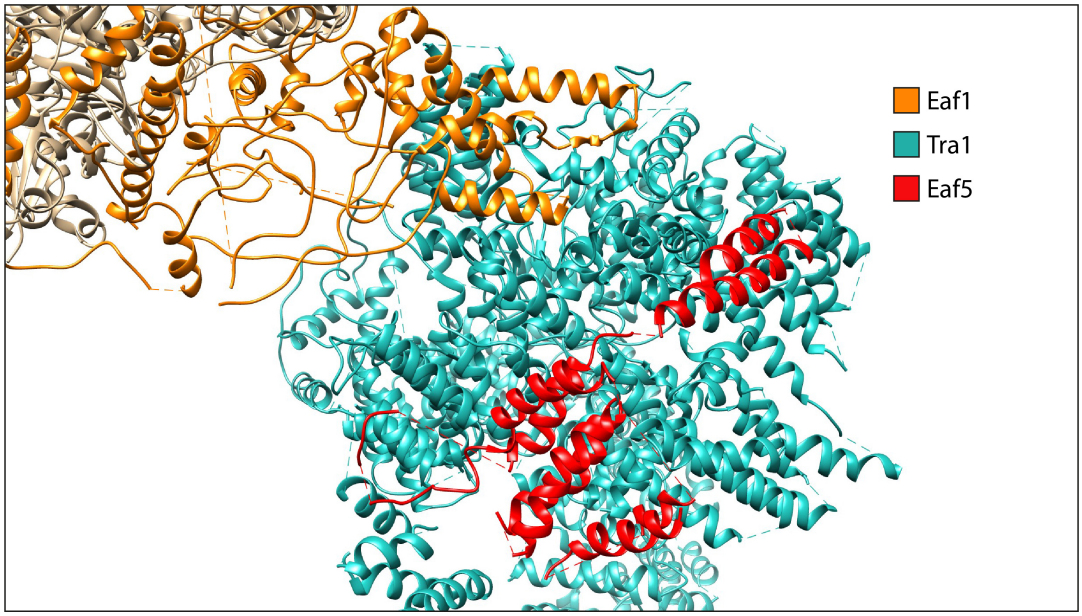

Figure 1.12: Part of the 4.7 Å structure of the TEEAA complex (PDB: 5Y81), in order to display Eaf5 interactions within the complex. The subunit forms a series of α-helices, extended over the large Tra1 subunit. Several regions within the structure remain unresolved, including contact between Eaf1 and Eaf5.

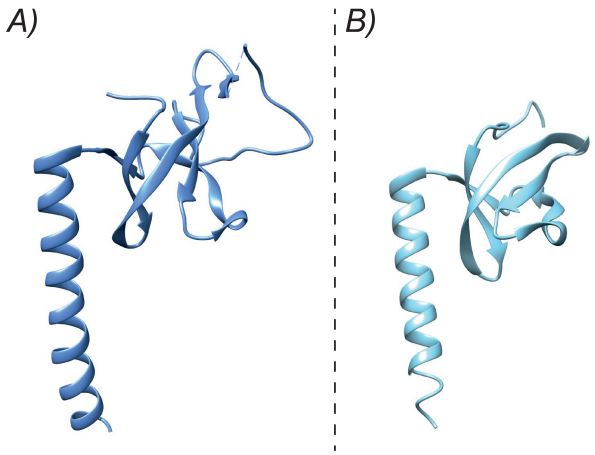

Figure 1.13: Ribbon representation of crystal structures of Eaf3 and the MRG15 chromo barrel domain. **A**) Crystal structure of Eaf3 chromo barrel domain (residues 1-124), at 2.5 Å resolution (PDB: 3E9G). **B**) Crystal structure of the MRG15 chromo barrel domain (residues 13-102) at 2.2 Å resolution (PDB: 2F5K).

Eaf3 also shares an MRG domain with human MRG15, towards its C-terminal. The domain acts as a binding interface with proteins, such as the rest of the Rpd3S complex and likely Eaf7 (Xie et al. 2015).

Eaf7 is negatively charged at physiological pH (isoelectric point ~5). Besides its negative charge and variable migration rates during SDS-PAGE analysis, little is known about the subunit's role in the trimer. Deletion studies by Rossetto et al. (2014), hint that the subunit acts as a scaffold for the other subunits to assemble around. Eaf7 shares ~30% sequence identity with human MRGBP, primarily in their N-termini. The C-terminal half of both proteins, have no apparent conserved domain and several disorder prediction algorithms suggest this region is highly disordered (**Appendix A**).

Human proteins, MRGBP and MRG15 form a dimer *in vivo*, that has been shown to directly interact with TIP60 and has been proposed to play a role in alternative splicing mechanisms (Gowher et al. 2012, Kirkwood et al. 2013). An NMR solution structure of the binding interface in the dimeric human complex, has been produced (**Figure 1.14**).

The Eaf3 ortholog (MRG15) had been shown to bind the MRGBP subunit (Eaf7 ortholog), using the same MRG domain that facilitates association with Rpd3S (Xie et al. 2015). This may enable the HAT and HDAC complexes to compete for the Eaf3 subunit, though Rossetto et al. (2014) purifications of Eaf3, find the subunit preferentially associates with the Rpd3S. The study also revealed that MRGBP associates with the MRG domain of MRG15, via a "minimal MRG-binding domain" (MBD), in its N-terminal. This interaction is maintained primarily by a series of hydrophobic interactions and a highly conserved FxLP motif (where x is any amino acid) within the MBD.

Eaf7 and MRGBP show highest levels of conservation around the MBD, though Eaf7 possesses a 31-residue insertion (residues 88-118) which is highly polar, as identified by Xie et al. (2015) (**Figure 1.15**).

Comparison of the hydropathicity of the two MBDs, using a grand average of hydropathicity index (GRAVY) score from ExPASy ProtParam, reveals that the Eaf7 MBD is significantly less hydrophobic than MRGBP MBD (-0.984 and 0.580, respectively, where positive values indicate hydrophobic sequences).

Eaf7 conserves 10 of the 28 key residues within the MRGBP MBD, that were shown to form hydrophobic interactions with the MRG domain. Whether this variation diminishes the strength of interaction between Eaf7 and Eaf3, relative to their human orthologs, has yet to be tested.

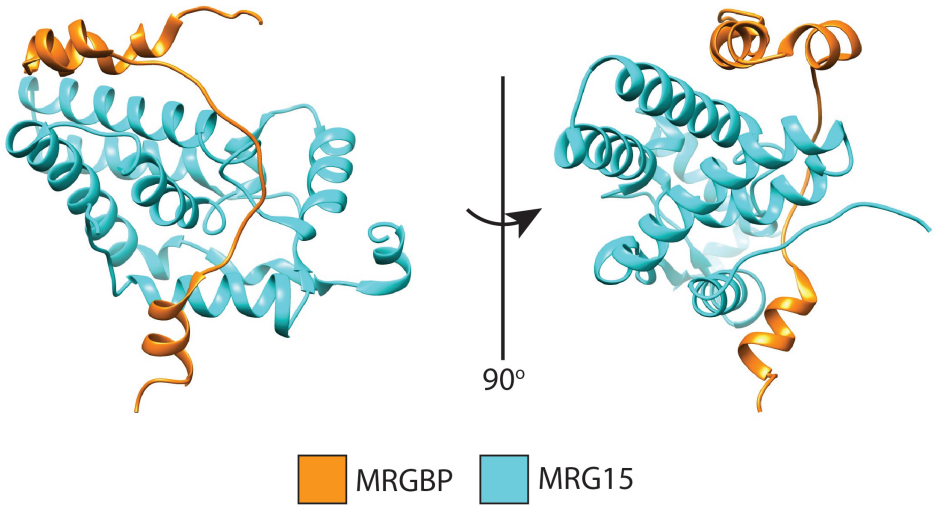

Figure 1.14: Ribbon representation of human MRG15 MRG domain (residue 155-323) associating with MRGBP MBD (residue 69-119), from an NMR study. The MRG domain is an essential binding interface and facilitates interactions between MRG15 and binding partners, such as MRGBP.

Eaf3 in comparison, has no major insertions in its MRG domain, with respect to the MRG15 protein. The two orthologs share ~33% sequence identity, though Eaf3 conserves 18 of the 33 residues in the MRG15 MRG domain, that were shown to form hydrophobic interactions with the MRGBP MBD. Several conserved residues in Eaf7 MBD and Eaf3 MRG domains, were shown to make contact in the human orthologous complex (**Figure 1.15**).

Whether these residues can interact, despite the relatively large insertion in Eaf7's MBD, has yet to be evaluated. A high-resolution structure of TINTIN would be of use, for a comparison of the yeast and human complexes. A summary of the conserved residues between the human and yeast subunits, that form points of contact in human MRG15/MRGBP, is listed in figure 1.15B.

A noticeable difference in the Eaf3 MRG domain, is the substitution of 5 of the 12 residues that make contact with the FxLP motif in the MBD. It is possible that differences in these positions, may also weaken interactions between the two domains. However, we cannot exclude the possibility that these substituted residues, make similar interactions to complete TINTIN assembly.

GRAVY scores for both Eaf3 and MRG15, suggest similar levels of hydrophobicity in both MRG domains (-0.18 and -0.32, respectively). It would be interesting to see whether Eaf3/Eaf7 interact via the MRG domain in the Eaf7 N-terminal, in the same way that the MRG15/MRGBP dimer is formed.

The interaction between Eaf5 and the rest of the complex is the least understood, though cross-linking mass spectrometry data, suggests the C-terminal of Eaf5 makes contact with Eaf7 N-terminal (Setiaputra et al. 2018). A summarising figure shows schematic representations of the TINTIN subunits, annotated to display conserved domains and highlights any regions that have been structurally characterised (**Figure 1.16**).

B)

Conserved MBD Residues	MRG Domain Interacting Residues
I77	E216, A219, G220, E223
W78	G220, Y224
L81	E216, L286, L287
Y85	K282, S283, L286
E94	T277, P278, L279
F105	Q232, L233, L234, Y235, R265, V268, R269
L107	L168, V169, W172, V268
P108	V268
E110	K157*
E114	K165*

Conserved residues

*Residues that make salt-bridges

FxLP Motif residues

Figure 1.15: A) Sequence comparison of the "minimal binding domain" (MBD) of the yeast Eaf7 subunit and the MBD of human ortholog MRGBP, identified by Xie et al. (2015). Residues that were shown to be essential for MRG15/MRGBP interactions, have been highlighted with an asterisk, these have been colour-coded, according to the type of interaction that occurs at that position. Any residue that is conserved, is marked with a red box. The 31-residue insertion in Eaf7, is displayed in the red dashed box beneath the sequence. **B**) Table listing conserved residues in the MRGBP and Eaf7 MBD, with the corresponding residues in the MRG15 MRG domain, that make direct points of contact based on NMR data (Xie et al. 2015). Residues are numbered according to their position in the human subunits. Residues that are listed in red, are conserved in the MRG domains of Eaf3 and MRG15.

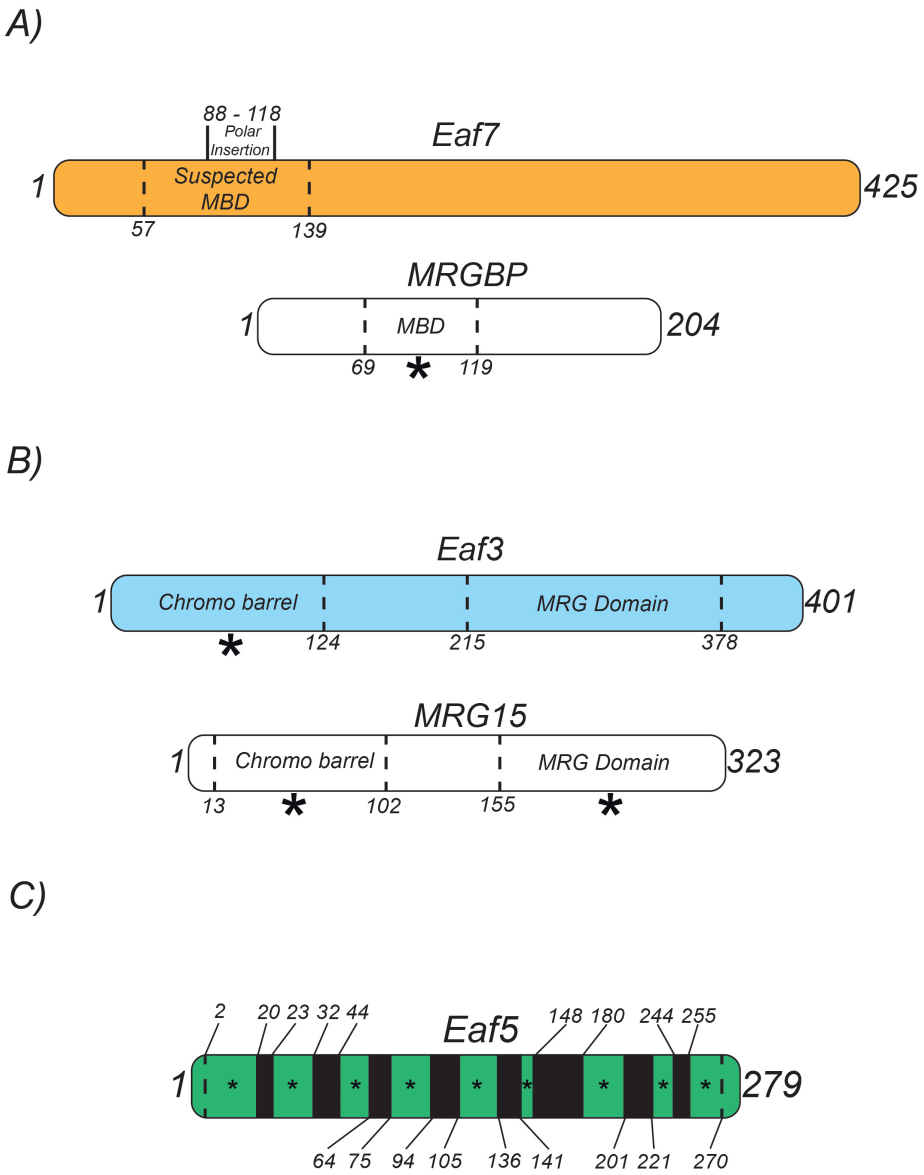

Figure 1.16: Schematic representation of the three TINTIN subunits, depicting conserved domains and their residue boundaries. Human orthologs MRGBP and MRG15 have also been displayed, for comparison with Eaf7 and Eaf3 respectively. Regions that have been structurally characterised have been marked with an asterisk. **A)** Eaf7 and the human ortholog, MRGBP. With the exception of an N-terminal region, that shares ~ 30% sequence identity with the MBD of MRGBP, Eaf7 shows no conserved domains within its sequence. **B)** Eaf3 and human ortholog, MRG15. The protein contains two distinct conserved regions, a chromo barrel domain (N-terminal) that has been resolved by X-ray crystallography and an MRG domain (C-terminal). **C)** Schematic of Eaf5. The protein is unique to yeast and has no known conserved domain within its sequence. Eaf5 has been partially characterised as part of a TEEAA "core" NuA4 structure. Unresolved regions are displayed as black bands and may correspond to flexible linkers between the short α-helices observed in the structure.

1.3.3 Biological Role

Eaf3, Eaf5 and Eaf7 mutants are sensitive to rapamycin, but not MMS, suggesting that the trimer is involved in NuA4's ribosome biogenesis activity, as opposed to its DNA damage repair functions. Furthermore, double mutants of these subunits and the Set2 methyltransferase, showed reduced growth when exposed to caffeine, mycophenolic acid (MPA) and 6-azauracil (6-AU); reagents implicated in nucleotide depletion and reduced transcription elongation rates (Reines et al. 2003, Rossetto et al. 2014).

The Set2 complex follows RNA Polymerase II during transcription elongation, associating with the phosphorylated CTD and methylating H3 in the wake of the transcribing complex (Schaft et al. 2003). This methylation (and phosphorylation of the Pol II CTD) leads to the stabilisation of the nucleosome, by recruiting histone deacetylases (HDAC) such as Rpd3S and remodelling complexes (Biswas et al. 2008, Govind et al. 2010, Heidemann et al. 2013, Corden. 2016). The function of this nucleosome stabilisation is to repress cryptic, spurious transcription within the body of the gene, after the DNA has been exposed by transcribing Pol II.

Set2 methylation is not essential for Rpd3S recruitment, as the Pol II CTD directly interacts with Rpd3S during transcription, but mutations in Set2 are often implicated in the transcription of intragenic material (Govind et al. 2010, Smolle et al. 2013). Interestingly, compensating mutations of TINTIN, can suppress intragenic transcription observed in Set2 mutants, suggesting that the trimer plays a role in nucleosome destabilisation (Rosetto et al. 2014).

Ginsburg et al. (2009) had already shown that NuA4 and SAGA appear to overlap in their positions along chromatin (UAS regions, upstream of transcription start sites) and were required for transcription activation and elongation by RNA polymerase II. However, while chromatin immunoprecipitation studies suggest that TINTIN associates with NuA4 over promoters, the subcomplex also associates with Rpd3S and Pol II over coding regions downstream (Schulze et al. 2011, Rossetto et al. 2014).

Moreover, pull-down of Pol II via tagged NuA4 subunits, is only possible when TINTIN is present and able to interact with the rest of NuA4. This may imply that the trimer is directly interacting with Pol II.

Pol II C-terminal domain (CTD) contains heptapeptide repeats with multiple phosphorylation sites. These sites are modified at different stages of the transcription process, allowing for the recruitment of specific factors to Pol II, at the correct time and position (Mitchell et al. 2008, Heidemann et al. 2013, Rossetto et al. 2014).

Ctk1 is responsible for the phosphorylation of the Pol II CTD, at Ser2 in the heptapeptide repeat. Deletion of the Ctk1 protein, disrupts the localisation of Eaf7 over coding regions (Rossetto et al. 2014). Eaf7 also preferentially pulls down Pol II phosphorylated at Ser2, a prevalent PTM in the late transcription elongation stage. Eaf5 appears to preferentially purify the Ser5 phosphorylated CTD, which is associated with earlier stages of transcription and this interaction is not dependent on the presence of Eaf7.

Taken into consideration, this data suggests that NuA4 components, particularly TINTIN subunits, are directly interacting with Pol II via its CTD, though no direct interaction has been observed. Given that these experiments were immunoprecipitation assays from cellular extracts, we cannot exclude the potential presence of additional factors, that may mediate interactions between TINTIN and Pol II.

Interestingly, ChIP analysis reveals that Eaf7 can be found over the coding regions of actively transcribed genes, when either Eaf5 or Eaf3 is deleted, albeit at a reduced signal. However, when Eaf3/Ctk1 double mutants are generated, this association is almost completely abolished. This suggests that TINTIN relies on the two interactions for correct recruitment; one with H3K36me3 (via Eaf3) and the other with the Pol II CTD (via Eaf5 and/or Eaf7).

Eaf7 is also shown to interact with the Spt16 subunit of the FACT complex, which is responsible for rearranging histones in preparation for transcribing RNA Pol II (Avvakumov et al. 2011). Rossetto et al. (2014) also showed that Eaf3/5/7 mutations (and mutations in Esa1), allowed for the expression of reporter genes, under the control of a cryptic transcription initiation site, provided chromatin remodelling complexes were available.

These same studies reveal synthetic lethality when Eaf5 and Eaf7 deletions are combined with SWR1 mutations, a remodelling complex responsible for incorporating the non-canonical H2A.Z into nucleosomes. SWR1 has been implicated in transcription elongation and splicing mechanisms (Neves et al. 2017), which further implicates TINTIN in the process of transcription elongation and potentially co-transcriptional splicing. The fact that the majority of genes that possess introns in yeast, are of ribosomal proteins, could explain why TINTIN mutants are sensitive to compounds that disrupt ribosome biogenesis, if the complex does indeed play a role in splicing.

Current research suggests that NuA4 is localised to transcription initiation sites, in order to carry out its HAT activity. The TINTIN trimer may then associate with elongating Pol II as it progresses downstream, recruiting remodelling complexes such as FACT and then the HDAC Rpd3S, to regulate NCP stability during transcription (**Figure 1.17**).

However, the exact mechanism of TINTIN recruitment remains unclear, as does the process by which it is released from the rest of the NuA4 complex, upon transcription initiation. Rossetto et al. (2014) also proposed that the trimer somehow promotes the processivity of RNA Pol II, though how exactly this might be achieved, remains unclear.

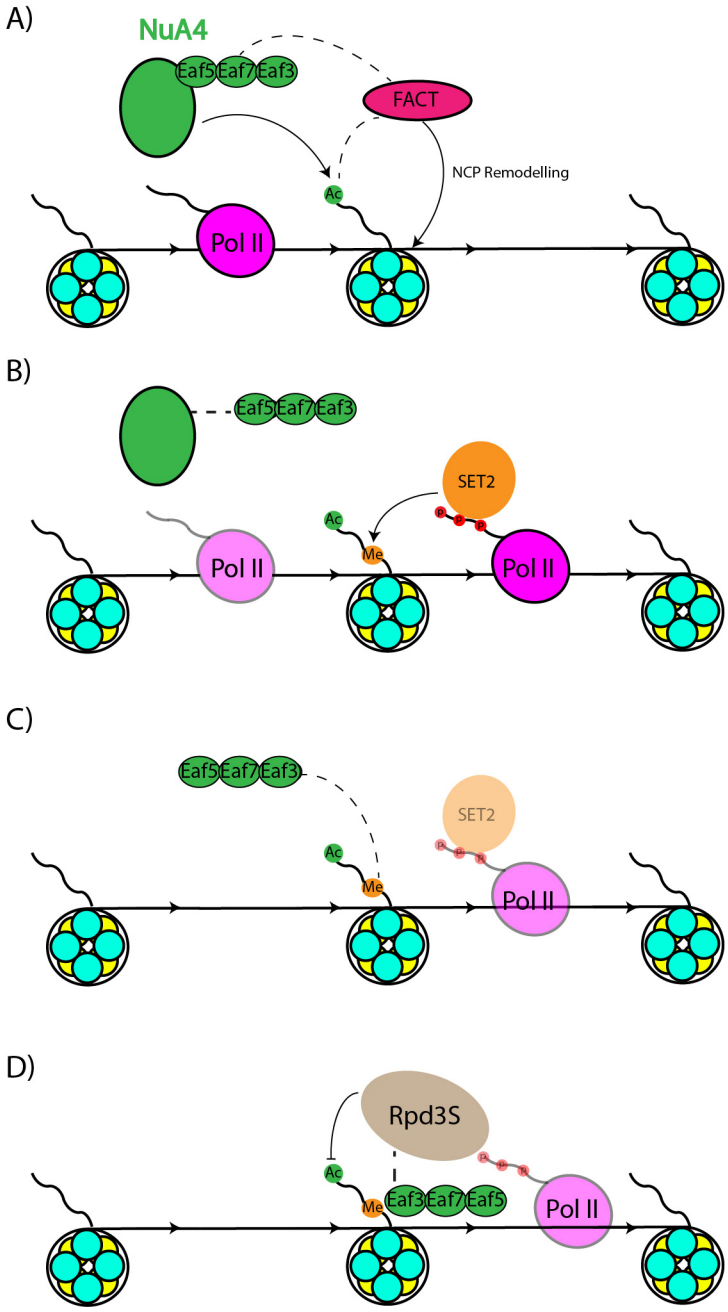

Figure 1.17: Schematic representation of the potential role of TINTIN, during transcription elongation. **A)** The NuA4 complex acetylates histones prior to the approach of transcribing RNA polymerase II. This acetylation and the Eaf7 subunit, recruits remodelling complexes such as FACT, which destabilises nucleosomes, facilitating the continued movement of the incoming Pol II. **B)** Pol II CTD is phosphorylated during transcription, recruiting the Set2 methyltransferase, which in turn methylates H3K36. **C)** H3K36me3 acts as a target for the Eaf3 chromodomain, which has been released (with Eaf7 and Eaf5) from NuA4. **D)** Recruitment of the Rpd3S complex by Eaf3 and Pol II CTD, which deacetylates histones in the wake of Pol II and helps stabilise the NCP.

1.3.4 Aims and Objectives

It is clear that TINTIN subunits play some important role in NCP stability and transcription elongation rates, though their exact mechanistic function remains obscure. If the trimer does indeed eject from the rest of NuA4, what triggers this transition?

The fact that NuA4 and Rpd3S act antagonistically *in vivo*, while sharing the Eaf3 subunit, raises the question if and how the subunit is passed from one complex to the other. TINTIN has no known catalytic activity and this relatively small complex may act as an intermediary, between the larger chromatin modifying complexes and Pol II. Additionally, the Eaf3 chromodomain has been shown to have a weak affinity for H3K36me3, but whether this affinity directs both TINTIN and Rpd3S recruitment to specific loci, is not clear. Importantly, the presentation of the Eaf3 chromodomain in TINTIN vs. Rpd3S might also alter its binding properties. We suspected that incorporation of Eaf3 into TINTIN might stabilise the chromodomain, modulating its affinity for H3K63me3 and providing additional binding to other features of the nucleosome.

One of the major objectives of this investigation, was to address these aspects of TINTIN function using structural, biophysical and biochemical techniques. Structural studies would be highly informative in elucidating the molecular mechanisms of the TINTIN complex and provide testable hypotheses for further study. The interactions of the TINTIN complex with Pol II and chromatin were to be verified biochemically, which would also evaluate the specificity of these interactions, ultimately aiming to confirm that the complexes preferentially associate during transcription elongation. This would be tested using *in vitro* pull-down assays, with purified sample and peptides mimicking chromatin or the CTD, at different stages of transcription.

Additionally, Eaf5 is unique to yeast, whereas the human orthologous complex to TINTIN, is composed of a dimer of MRG15 (Eaf3 ortholog) and MRGBP (Eaf7 ortholog), hence the function of Eaf5 is ultimately unclear. Some significant sequence differences in the minimum MRG binding domain of Eaf7 (compared to the MBD of MRGBP), were proposed to weaken interactions with the MRG domain of Eaf3. It was also speculated that the additional Eaf5 subunit, plays a role in stabilising the structure, as well as bridging the complex to the rest of NuA4.

The structure of the TEEAA complex, suggests that the interaction between Eaf5 and the rest of TINTIN is weak. We therefore aimed to evaluate the strength of this interaction, by purification and native mass spectrometry, followed by collision induced dissociation.

Finally, over 50% of the Eaf7 subunit is predicted to be disordered from its C-terminal, despite the subunit apparently playing a scaffolding role between Eaf5 and Eaf3 (Rossetto et al. 2014) (**Appendix A**). Another objective of this experiment was to identify the essential regions within Eaf7 (and other TINTIN subunits), for complex assembly and stability. We theorised that the C-terminal of Eaf7 was dispensable for complex assembly.

To confirm this theory and to identify the boundaries of interaction interfaces between subunits, a series of truncated TINTIN constructs would be generated. These truncations would also be used in Pol II CTD pull-down studies, to help identify features responsible for this interaction. This would be carried out in parallel to structural analysis, with techniques such as X-ray crystallography and SAXS. The aim being to visualise the architecture of the complex and ideally, identify binding regions in the complex for NuA4, Rpd3S and Pol II interaction.

Chapter 2

Materials and Methods

2.1 Construct Design and Cloning

Throughout the project, several TINTIN constructs, including the full-length complex and constructs with truncated subunits were designed. A pRSFDuet-1 plasmid containing subunits Eaf1, Eaf5, Eaf3 and Eaf7 with a C-terminal 3C protease site, followed by a Twin-Strep tag, had already been generated by Dr Alan Cheung. The Eaf5/7/3 sequences (and cleavage site + tag) were cloned into a pET-28a plasmid (**Figure 2.1**).

All plasmids were designed using the open source software ApE (A Plasmid Editor), with primers designed using the same software. Primers were designed to contain 3' overhangs of at least 20 nt for Gibson assembly reactions (Gibson et al. 2009), these overhangs would contain complementary sequences between PCR products we wished to ligate. Annealing temperatures of primers were calculated with NEB Tm calculator.

50μL PCR reactions were carried out using the New England Biolabs Q5 hot start high-fidelity DNA polymerase. The reaction scheme followed the volumes and concentrations recommended by NEB, with the addition of 5% DMSO as listed in **Table 2.1.**

Components	Stock Concentration	Volume (μL)	Final Concentration
Q5 Reaction Buffer	5X	10	1X
dNTPs	10 mM	1	200 μM
F/R Primer	10 μM	2.5	0.5 μM
DNA Template	1 ng/μL	1	20 ng/μL
Q5 Polymerase	2 units/μL	0.5	0.02 U/μL
DMSO	100%	2.5	5%
H_2O	-	30μL	-

Table 2.1: Summary of components used for 50 μL PCR reactions using the NEB hot start Q5 high-fidelity DNA polymerase. Mili-Q ultrapure water was used for all reactions.

The PCR reaction followed the same general scheme. Beginning with denaturation at 98 °C (35 seconds) followed by 25 cycles of: 98 °C (10 seconds), annealing (primer-dependent temperature) (30 seconds), elongation at 72 °C (30 seconds/kb). This was completed with a final extension step at 72 °C for 2 minutes.

The success of PCR reactions was determined by gel electrophoresis. Samples were first mixed with 10 μL of 6X DNA loading dye, 50 μL of the resulting mixture was then loaded onto the gel (1X TAE, 1% agarose, 1X Invitrogen SYBR safe dye) and run for 60 minutes at 100 V.

If bands were revealed by UV detector, they were excised and processed by a QIAquick gel extraction kit (Qiagen) and eluted with 30 μL elution buffer. The concentration of these PCR products was determined by NanoDrop spectrophotometer. Fragments to be annealed would be mixed at a ratio of 0.04:0.04 pmols, and incubated at 50 °C for 45 minutes in equal volume with a 2X NEBuilder HiFi DNA assembly master mix, a commercially available Gibson reaction kit.

The resulting Gibson assembly reactions were then added directly to 50 μL aliquots XL10-Gold competent cells. These were then incubated on ice for 30 minutes before heat shock (42 °C for 45 seconds), followed by the addition of 500 μL of SOC rich media and further incubation at 37 °C for 1 hour.

Cells were then pelleted (5 krpm for 2 minutes) and then streaked onto LB agar plates containing 50 μg/mL kanamycin. Plates were then incubated at 37 °C for 16 hours. If colonies were observed after this incubation, they were picked and incubated in 5 mL of LB (including μg/mL kanamycin) for an additional 16 hours.

Assuming growth was observed, plasmids were purified from these overnight incubations, using the QIAprep Miniprep kit (Qiagen). Samples of these purified products were sent to GATC for sequencing, to confirm the successful generation of constructs.

2.1.1 Expression Vectors

The majority of constructs cloned used the pET-28a vector, which confers kanamycin resistance (**Figure 2.1**). The backbone contains a T7 promoter under the control of the lac operon (Dubendorfft et al. 1991).

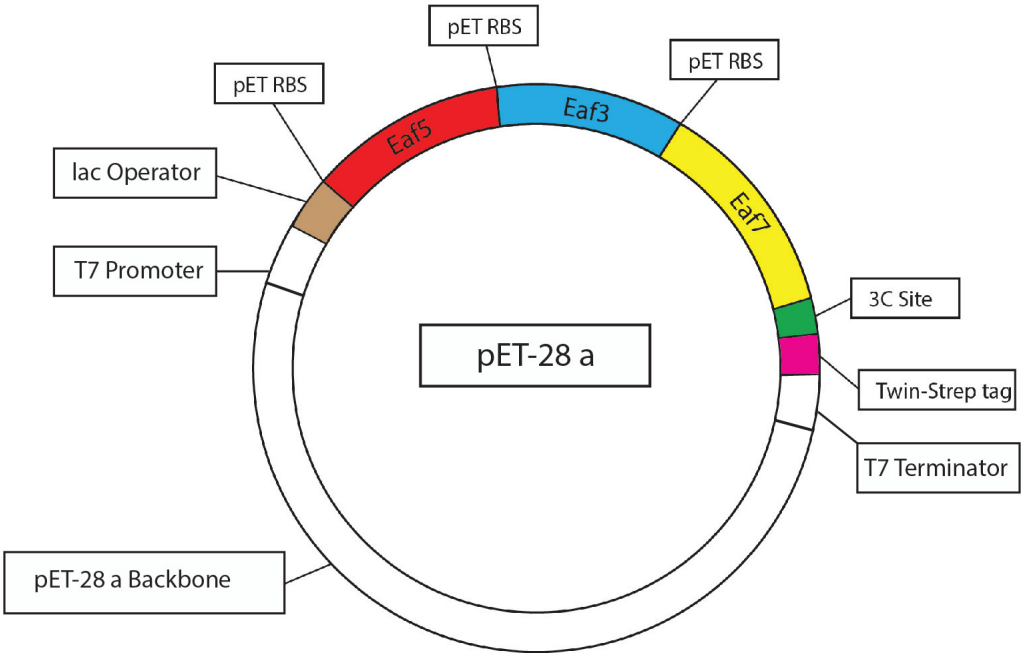

Figure 2.1: Schematic of the pET-28a vector and representative of the full-length TINTIN construct that acted as the foundation for this project. The low-copy number plasmid confers kanamycin resistance to transformed cells, and contains a T7 promoter under the influence of a *lac* operon.

2.1.2 Constructs Generated

A series of truncated constructs were derived from the full-length TINTIN complex, subjected to test expressions and purified for the purpose of this project (**Figure 2.2**). All constructs included a C-terminal 3C protease site, followed by a Twin-Strep tag.

Truncation boundaries were informed by limited proteolysis and native mass spectrometry results, later discussed in this chapter and results.

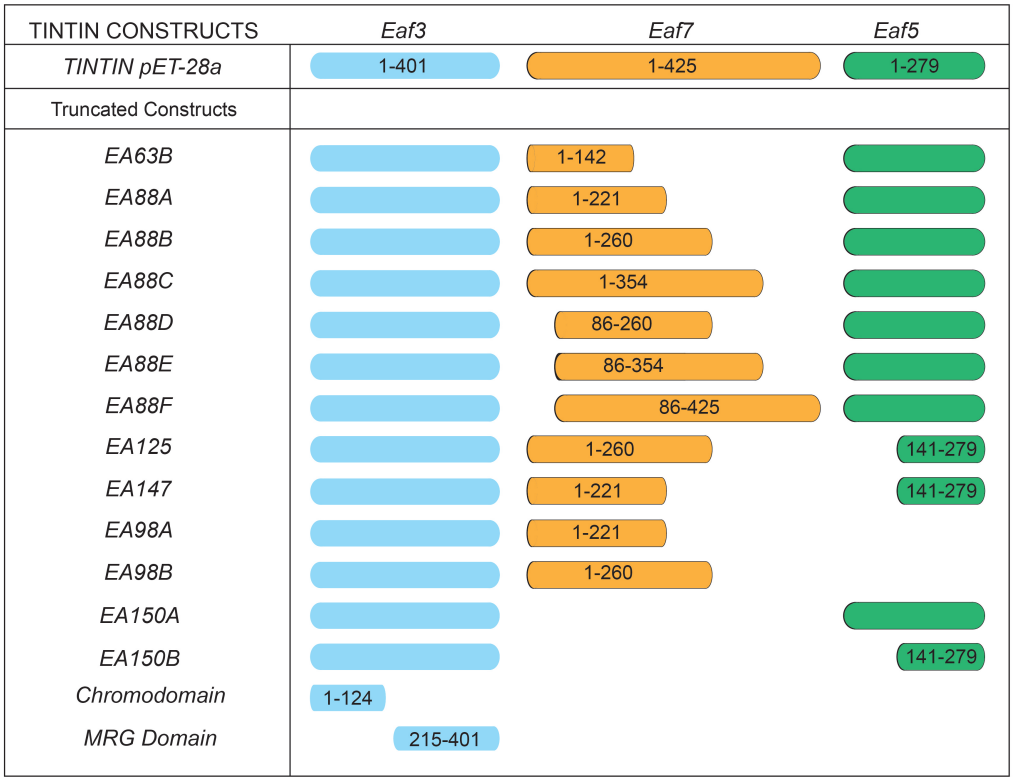

Figure 2.2: List of TINTIN associated constructs cloned and subjected to test expressions throughout this project. Schematics of each subunit are included and the boundaries of any truncation are listed within the schematic of the subunit in question. The length of Eaf7 subunits in the figure does not include the C-terminal 3C protease site and affinity tag.

2.2 Protein Expression

Rosetta 2 competent cells were used to express the TINTIN-pET-28a construct and had been transformed identically to the XL10-Gold competent cells in the cloning process. Preliminary (small-scale) expressions tested the efficacy of 2 types of media across 4 different expression protocols in Rosetta 2 cells (**Table 2.2**). The media included were LB and the autoinduction (AI) media described by Studier (2005); ZYM-5052 (referred to as AI media for this report).

AI media based expression, relies on growth of cells in a mixture of glucose, α-lactose and glycerol as sources of carbon. Cells grow in this media readily, while the T7 lac promoter is repressed in the presence of glucose. Upon depletion of glucose (which is preferentially metabolised before other carbon sources), cells have grown to higher OD (than they would in media such as LB), before the α-lactose driven induction begins.

The aim of this method is to increase overall levels of protein expression by way of higher cell count, with the added advantage of allowing for cells to grow before potentially toxic recombinant proteins are expressed. Formulation of this media is summarised in Table 2.3.

Transformed cells were used to inoculate 5 mL LB pre-cultures, including 50 µg/mL kanamycin and 34 µg/mL chloramphenicol (resistance conferred by the Rosetta 2 strain). The resulting overnight culture was used to inoculate AI, TB or LB containing flasks (at a ratio of 1 mL pre-culture to 500 mL of expression media).

For upscaled AI media expressions, 12 L of AI media was used. This was increased up to 24 L for LB/TB expressions, depending on the construct being expressed. Cells were harvested at 4000 rcf, with the supernatant removed and cell pellets flash frozen in liquid nitrogen for storage.

Condition	Pre-Induction	Induction OD	[IPTG] (mM)	Post-Induction
AI1	37 °C	0.5	-	18 °C(24 hours)
AI2	37 °C	0.5	-	37 °C(16 hours)
LB1	37 °C	0.5	0.2	37 °C(4 hours)
LB2	37 °C	0.5	0.2	18 °C(18 hours)

Table 2.2: Summary of test expression schemes used in the production of TINTIN. All cells were shaking at 200 rpm throughout their expressions. Protocols in the table are displayed as: pre-induction temperature, OD cells reach before induction, IPTG induction concentration and the post-induction conditions cells are incubated in. Conditions listed as AI1 and AI2 used autoinduction media and therefore did not require IPTG driven induction.

Stocks	Compounds	Concentration
50X M	Na_2HPO_4	1.25 M
	KH_2PO_4	1.25 M
	NH_4Cl	2.5 M
	Na_2SO_4	0.25 M
$MgSO_4$	-	1.0 M
50X 5052	Glycerol	25% (w/v)
	Glucose	2.5% (w/v)
	α-lactose	10% (w/v)
1000X Trace Metals	$FeCl_2.6H_2O$	0.1 M
	$CaCl_2$	1.0 M
	$MnCl_2.4H_2O$	1.0 M
	$ZnSO_4.7H_2O$	1.0 M
	$CoCl_2.6H_2O$	0.2 M
	$CuCl_2.2H_2O$	0.1 M
	$NiCl_2.6H_2O$	0.2 M
	$Na_2MoO_4.2H_2O$	0.1 M
	$Na_2SeO_3.5H_2O$	0.1 M
	H_3Bo_3	0.1 M

Table 2.3: Summary of stock buffer compositions used for formulation of ZYM-5052 media designed by Studier (2005). Stocks were stored separately and diluted with ZY media (1% tryptone (w/v) and 0.5% yeast extract (w/v)) to prepare 12 X 1 L media in 2 L conical flasks. M and 5052 media are diluted to 1X, while trace metals are diluted to 0.2X and $MgSO_4$ is diluted to 2 mM.

2.3 TINTIN Properties and Purification Design

Prior to purifications attempts, the ExPASy ProtParam tool was used to approximate some of the properties of Eaf5, Eaf7 and Eaf3.

Eaf7 (including a 3C protease site followed by the Twin-Strep tag at the C-terminal position), has a theoretical molecular weight of ~53 kDa and an isoelectric point of 5.19. Eaf3 has a theoretical MW of ~45 kDa and isoelectric point of 8.5. Eaf5 has a theoretical MW of 31 kDa and isoelectric point of 8.03.

The combined theoretical molecular weight of the complex is ~130 kDa with an averaged isoelectric point of 5.95. It was noted that the literature suggested that Eaf7 may not migrate across an SDS-PAGE gel at a rate expected of a protein of equivalent molecular weight, this would be verified by our own SDS-PAGE gels and mass spectrometry.

2.4 Purification Summary

All purifications followed the same general scheme:

1. Lysis via sonication.
2. Clarification of cell-lysate by centrifugation.
3. Strep-Tactin purification.
4. Ion-exchange chromatography (IEX).
5. Size-exclusion chromatography (SEC).
6. Final spin concentration of sample.

2.4.1 Purification Buffer

All purifications used the same stock buffer A (100 mM Tris/HCl (pH 8.0), 1 mM EDTA, 10 μM $ZnCl_2$, 150 mM NaCl), which is recommended for Strep-Tactin based purifications. Modifications to this formula are listed where appropriate. Thermofluor data (**Appendix B**) suggests that TINTIN is stabilised in lower salt concentrations and NaCl concentration was reduced to 50 mM in the final, optimised purifications.

2.4.2 Lysis and Clarification

Cells pellets were resuspended in buffer A (with additional 0.1% NP40 + 2X Sigma-Aldritch protease inhibitor cocktail tablets). After resuspension, cells were lysed by sonication at 75% amplitude at 3 x 10 minute intervals. The resulting lysate was subjected to centrifugation at 17 krpm for 1 hour in a Sorval SLA-1500, the supernatant from this was then applied to Strep-Tactin resin.

2.4.3 Twin-Strep Tag System

The recombinant tag employed in purification of the TINTIN trimer, was a Twin-Strep tag designed to bind to the engineered Streptavidin: Strep-Tactin. Twin-Strep tags contain two copies of the conventional Strep-tag (WSHPQFE), which are separated by a short spacer region (serine, alanine). This allows for increased levels of binding and specificity of binding between the tagged protein and Strep-Tactin resin.

Elution from Strep-Tactin resin and Strep-TactinXT (which is designed to bind Twin-Strep tagged proteins as opposed to a single Strep-tag) is facilitated by the application of 50 mM desthiobiotin and biotin respectively, the required volume of elution buffer (Buffer A + 0.1% NP40 + 50 mM biotin/desthiobiotin + 2X Sigma-Aldritch protease inhibitor cocktail tablets) was subject to optimisation and is discussed in the results chapter 3.

2.4.4 Chromatography System

An ÄKTA Pure chromatography system was used for ion-exchange and size-exclusion chromatography steps. Method creation, data collection and analysis was carried out using the UNICORN 7 software. Prior to use, all columns were washed with 1.5 column volumes (CV) of degassed H_2O and equilibrated in 1.5 CV of buffer.

2.4.5 Ion-Exchange Chromatography

Prior to any use of the ÄKTA Pure system and its high-resolution columns, samples were filtered with Sartorius Minisart NML syringe filters, to remove precipitate which could potentially damage the columns. The two columns used in this step include a Mono Q 5/50 (GL) and Mono S 5/50 (GL) for anion and cation exchange respectively.

Due to the large volume of Strep-purified protein, columns were loaded with TINTIN using the systems' sample pump. The system was eventually programmed so that the flow-through of the Mono Q column was immediately loaded onto the Mono S, followed by collection. After sample application, columns were washed with 10 CV buffer A before an increasing percentage buffer B (buffer A + 2 M NaCl) over 20 CV. A final high-salt wash of 100% buffer B was applied to remove any proteins still bound to the columns. Samples were eluted into 0.5 mL fractions.

2.4.6 Size-Exclusion Chromatography

SEC steps also use buffer A. Protein samples to be loaded onto these columns, were spun down (13 krpm for 10 minutes at 4 °C) before being injected onto the ÄKTA Pure loop. The volume of the loop used, ranged from 0.5 mL to 5 mL, depending on the column used and volume of TINTIN sample being loaded. All loops were washed with 2X their volume of buffer A to apply to the SEC columns.

Either the Superdex 200 increase 10/300 GL or the HiLoad 16/600 Superdex 200 pg were used in purifications, both are able to resolve proteins within the MW range of 10,000 to 600,000. Either column was washed with 1.5 CV of buffer A to elute the protein. Samples were eluted into 2 mL fractions.

Calibration curves were made for these columns (**Appendix C**) by Dr Alan Cheung, which allowed for MW approximations of experimental proteins, based on their elution profiles.

2.4.7 Protein Concentration

TINTIN constructs were concentrated for the purpose of crystallisation. This proved to be a major bottleneck in the final yield of the protein and as a result, multiple spin concentrators (varying in MW cut-off and membrane type) were tested for their efficacy (**Table 2.4**).

Concentrator	Membrane Type	Mw Cut-Off (kDa)
Amicon Centrifugal Filter	Regenerated Cellulose	10
Amicon Centrifugal Filter	Regenerated Cellulose	30
Amicon Centrifugal Filter	PES	10
Amicon Centrifugal Filter	PES	30
Centriprep Centrifugal Filters	Regenerated Cellulose	10

Table 2.4: Summary of concentrating equipment used to prepare the purified TINTIN complex for crystallisation trials.

After multiple optimisations of this concentration process (data not included), it was found that the devices that resulted in minimal loss of protein were the Centriprep centrifugal filter (10 kDa cut-off) and the amicon centrifugal filter (10 kDa cut-off, regenerated cellulose). 5-10% glycerol was also included in the final purification buffer, in an attempt to reduce protein loss, but was found to be ineffective (data not included).

2.4.8 Quantification

Concentration of the sample was generally determined using a NanoDrop spectrophotometer, though for post-Strep purified protein (which contained the detergent NP40), concentration was quantified by Bradford assay using a commercially available quick-start Bradford assay kit and protocol (Bio-Rad).

2.4.9 SDS-PAGE

Purified samples were diluted with a 6X SDS-PAGE loading buffer (360 mM Tris pH 6.8, 10% SDS, 50% glycerol, 0.03% bromophenol blue, 600 mM DTT). Diluted samples were boiled for 2 minutes before loading onto a commercially available Bolt Bis-Tris gel, immersed in MES buffer. Gels were run at 200 V for 45 minutes and stained using InstantBlue Coomassie stain.

2.4.10 Mass Spectrometry

After SDS-PAGE analysis of the final purified product, bands that were proposed to correspond to TINTIN subunits, were excised and placed into 1.5 mL Eppendorf tubes and sent to St. Andrews Mass Spectrometry service, for protein identification by MALDI MS & MSMS (Egelhofer et al. 2002). This was also employed when suspected degradation products were observed in gels.

2.5 Crystallisation Trials

The crystallisation of protein samples is an essential step in the production of high-resolution structures by X-ray crystallography (McPherson and Gavira 2014). The process of crystallisation is complex, depending on several factors including (but not limited too): sample concentration, salt concentration, pH, temperature, sample homogeneity and structural flexibility (**Figure 2.3**).

We aimed to crystallise the full-length TINTIN complex for X-ray crystallography purposes, using a range of commercial screens, sample concentrations, temperatures and adjustments to buffer. Additionally, truncations were generated in the TINTIN complex (primarily of the Eaf7 subunit) in order to generate a more stable core complex, potentially removing flexible regions that were inhibiting the crystallisation process.

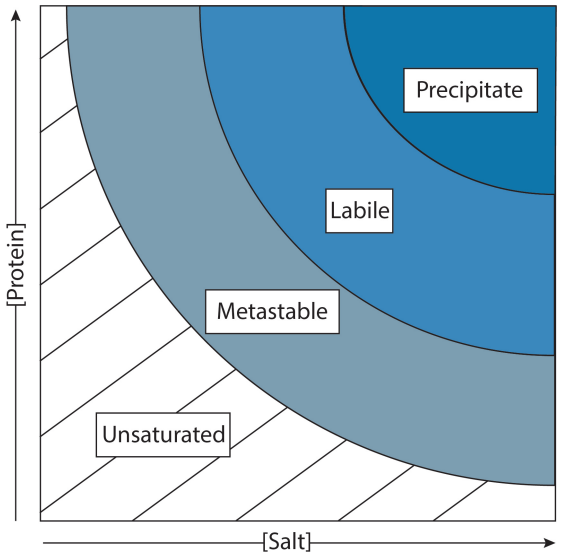

Figure 2.3: Summary of optimal crystallisation conditions, which are dependent on salt and protein concentration, but also pH, temperature, homogeneity and protein flexibility. Crystallisation is a state of supersaturation, sample and solvent concentration must be balanced to ensure this supersaturation is reached without precipitating the protein. Crystals (which can form in metastable and labile regions) must nucleate and grow to significant dimensions for X-ray analysis, this can only occur in the theoretical labile region.

2.5.1 Commercial Screens

Crystallisation screens were carried out in parallel with the TINTIN purification optimisation (**Table 2.5 and 2.6**). The concentration of TINTIN for each round of crystallisation attempts, reflects the maximum yield of the trimer achievable at that point in time. As a result, increasing concentrations of TINTIN were used with each attempt.

Sparse Matrix Screen	Concentration of TINTIN (mg/mL)			
	4.2	7.0	12.4	10.0 (4 °C)
JCSG-plus	X	X	X	X
Pact	X	X	-	X
SaltRX HT	X	-	X	X
Index	X	X	X	X
Structure Screen 1	X	X	X	X
Midas	X	-	-	-
ProPlex	X	X	X	-
Morpheus HT-96	-	X	X	X

Table 2.5: Summary of crystallisation attempts of full-length construct. Each commercially available sparse matrix screen is given an 'X' if it was used for a corresponding TINTIN preparation. Protein sample and buffers from each screen were mixed 1:1 (v/v) by Mosquito crystal for accuracy and reproducibility. The resulting plates were stored at 16 °C in a Rigaku Minstrel DT imager and gallery DT plate hotel. Screens that were stored at 4 °C, were periodically checked for crystal formation by light microscope.

Screens	EA88A		EA88B	
	Concentration (mg/mL)			
	8*	11	5	6
Index	X	X	X	X
JCSG	X	X	X	X
Pact	X	X	X	X
Morpheus	X	X	X	X
Structure	X	X	X	X

Table 2.6: Summary of crystallisation trials of truncated TINTIN constructs. Screens were incubated at 16 °C in the Rigaku Minstrel DT imager and gallery DT plate hotel and periodically imaged for the presence of crystals. The concentration marked with an asterisk was tested at both 16 °C and 4 °C.

TINTIN truncations (EA88A and EA88B) were also purified and subjected to crystallization trials using the Index, JCSG, Pact, Morpheus and Structure commercial screens. Commercially available, sparse matrix screens were used to cover a wide range of crystallisation conditions. TINTIN samples were mixed with each condition (150 nL sample + 150 nL buffer) by Mosquito crystal into MRC 2-well vapour diffusion plates.

These were stored at 16 °C in a Rigaku Minstrel DT imager and gallery DT plate hotel. Each plate was immediately imaged upon loading onto the hotel platform, allowing for detection of any immediate crystal formation. Samples were then periodically examined by the device and saved on a database for review.

2.5.2 Custom Hanging Drop Screens

If suspected TINTIN crystals were identified in any of the commercial screens, the buffer that corresponded to the well in question, was scrutinised and used to design multiple hanging-drop screens, in an effort to improve or replicate the crystallization process.

Qiagen hanging-drop plates were used, containing varying concentrations of purified protein +/- 10% and varying concentrations of precipitants. Wells contained 500 μL of a precipitant mix, with hanging-drops containing 1 μL purified protein and 1 μL precipitant mix. This maintained the ratio of 1:1 (volume/volume) used in the test commercial screens.

2.6 Differential Scanning Fluorimetry Assays

Differential scanning fluorimetry (Thermofluor) assays were conducted on purified full-length TINTIN (2.48 mg/mL), across a range of buffer conditions obtained from the RUBIC buffer screen (Molecular Dimensions). The aim was to identify stabilising buffer conditions. (**Appendix B**).

25 µL reactions were carried out in a MyIQ real-time PCR instrument using the Sypro Orange dye. The raw fluorescence signal vs temperature data was converted to -dRFU/dT vs T, a derivative plot, which displays changes in fluorescence vs temperature, allowing for easier determination of Tm (melting temperature).

Buffer conditions were from the commercially available RUBIC buffer screen. A positive control of 12 μL 2 mg/mL lysozyme + 12 μL buffer was used. A negative control of 24 μL buffer was also included. Sypro Orange absorbs light of wavelength 470 nm and emits at 569 nm.

Reactions were carried out in PCR-microplates, with each well containing 21 μL of RUBIC buffer, followed by the addition of 2 μL protein and finally 2 μL Sypro Orange at a final concentration of 5X from the original stock (5000X).

Plates were sealed with ClearVue seals, before being placed in the RT-PCR device, pre-equilibrated to 5 °C. The device was programmed to gradually increase the temperature from 5 °C to 95 °C (at a rate of 1 °C/minute). Filters used were compatible with the maximum excitation/emission wavelengths of Sypro Orange.

Experiments were carried out in duplicate and averaged, with melting temperatures of the TINTIN complex in each of the RUBIC buffers compared to determine conditions of higher stability.

2.7 Limited Proteolysis

The JBS Floppy-Choppy limited proteolysis kit was used for limited proteolysis experiments, with the objective of removing regions with high levels of flexibility from the TINTIN complex.

The proteases included in these experiments included: α-chymotrypsin, trypsin, subtilisin and papain. Initial attempts at proteolysis followed a manual recommended ratio of TINTIN:protease (1 μg:1 ng) in a reaction volume of 40 μL (containing TINTIN purification buffer). Samples were incubated for 30 minutes at 22 °C and quenched with 8 μL of 6X SDS loading dye. The extent of proteolysis was then determined by SDS-PAGE by loading 20 μL of this reaction mixture.

Optimised reactions contained 2.8 μg of TINTIN for better visualisation on the SDS-PAGE gel. Protease quantities were optimised for progressive levels of proteolysis (**Table 2.7**). Temperature was maintained at 22 °C and reactions were carried out in triplicate, quenching at 10 minute intervals (**Table 2.7**).

If bands observed on the SDS-PAGE gel were intense, discreet from other bands (not part of a smear) and reproducible, these were excised and stored in 1.5 mL Eppendorf tubes. Stored samples were sent to St. Andrews mass spectrometry and proteomics facility for sample identification. This data was used to determine boundaries for cloning of truncated TINTIN constructs.

Protease	Protease Quantity (ng)	Incubation Time (min)
Papain	140	25
	140	35
	140	45
α-Chymotrypsin	140	10
	140	20
	140	30
Trypsin	2.8	10
	2.8	20
	2.8	30
Subtilisin	2.8	10
	2.8	20
	2.8	30

Table 2.7: Summary of optimised limited proteolysis experiments on the full-length TINTIN complex with the JBS Floppy-Choppy kit. 2.8 μg of TINTIN was mixed with varying quantities of protease (in a 40 μL reaction) and quenched at varying time points with 8 μL of 6X SDS loading dye. All reactions were carried out at 22 °C, with temperatures maintained using a heat block.

2.8 Disorder Prediction

TINTIN subunits were submitted to the MobiDB database (Piovesan et al. 2018), to identify regions of flexibility and disorder within each of the subunits. This would also be used to identify targets for truncation.

The database provides multiple outputs from various software packages, then overlays and averages these results, so that users can have more confidence in disorder prediction (**Appendix A**).

2.9 TINTIN Interaction/Pull-down Assays

As proposed by Rossetto et al. (2014), the TINTIN trimer has been predicted to bind to the phosphorylated CTD of RNA Polymerase II, though this interaction had yet to be directly tested.

As a result, purified TINTIN constructs EA88A, EA88B and the Eaf3 chromodomain (the long form described in Sun et al. 2008), were subjected to pull-down assays (Wysocka, 2006), using peptides that mimicked the Pol II CTD. These peptides contained phosphorylation patterns representative of the CTD at different stages of transcription elongation, as well as various methylated histone peptides (mimicking the H3 N-terminal tail), that potentially interact with the chromodomain of Eaf3.

2.9.1 Peptides

Biotinylated peptides were designed to bind Strep-Tactin dynabeads. Once fixed to the resin, the peptide would be exposed to the protein sample, followed by multiple wash steps. If interaction between peptide and protein occurred, then SDS-PAGE analysis of the resin, would reveal the presence of the applied protein.

Histone and CTD peptides were chemically synthesised (Sigma-Aldrich and Alta Bioscience respectively) and provided in a lyophilized form. All peptides were dissolved to 0.5 mg/mL in commercially available PBS buffer (Sigma-Aldrich), with CTD-peptides being dissolved in PBS + 20% DMSO (**Table 2.8.**). CTD peptides were designed to replicate 2 consecutive heptad repeats, with post-translational modifications present on both of the repeats.

	Peptides	% DMSO (v/v)
Histone (N-terminal)	H3 (21-44)	-
	H3K36me1	-
	H3K36me2	-
	H3K36me3	-
	H3K4me3	-
	H4 (1-21)	-
Pol II CTD	CTD-blank	20%
	Y1p	20%
	S2p	20%
	S5p	20%
	Y1p/S2p	20%

Table 2.8: Summary of biotinylated peptides used for TINTIN pull-down assays. All peptides were diluted to 0.5 mg/mL in PBS buffer, though Pol II CTD peptides were diluted with additional 20% DMSO (v/v). Post-translational modifications are listed with the corresponding histone peptide, for example H3K36me1 = lysine 36 of H3, with a single methyl group. CTD peptides are listed with as the post translational modification pattern they correspond to, for example; Y1p = phosphorylated tyrosine 1.

2.9.2 Strep-tag cleavage

The C-terminal Twin-Strep tag used in all constructs was removed with a GST-tagged 3C protease, to prevent binding to the Strep-Tactin resin used in the pull-down assay. 98 μL of purified protein (~4 mg/mL) was cleaved with 10 μL 3C protease (1.5 mg/mL) and 2.2 μL DTT (50 mM), incubating for 1 hour at room temperature.

After incubation, protease was removed with GST resin (2 μL samples per 1 μL of resin), incubating the sample and resin for 15 minutes, before pelleting with a tabletop centrifuge at maximum speed for 2 minutes.

The supernatant was then applied to Dynabead (Strep-Tactin) resin and the process was repeated to remove the cleaved tag. The final protein sample, cleared of the C-terminal Twin-Strep tag and protease, was diluted with purification buffer to 1 mg/mL.

2.9.3 Pull-down Assay

The pull-down assay is summarised in Table 2.9. 40 μL Strep-Tactin Dynabeads were used to fix the biotinylated peptides. Controls (resin + protein) were also included to evaluate the extent of non-specific binding between TINTIN and resin.

Component	[Stock](μg/mL)	Volume (μL)	Incubation (seconds)	Wash Buffer
Dynabeads	-	-	-	PBS-T
Peptide	0.5	15	120	PBS-T
Protein	1	15	120	Buffer A-T

Table 2.9: Summary of pull-down reaction scheme. Reaction components are listed sequentially, with wash buffer (applied in 3 x 500 μL batches) applied after the addition of each component. All buffers included 0.1% (v/v) triton X-100 (denoted "-T"). Reactions were incubated at 25 °C while shaking (1200 rpm).

Assays were carried out at 25 textcelsius while shaking (1200 rpm) which was maintained by heat block. The addition of each component was followed by a wash step of 3 x 500 μL wash buffer (either PBS or purification buffer A with 0.1% (v/v) triton X-100). Triton X-100 was included to reduce non-specific binding.

The success of each assay was determined by SDS-PAGE analysis. After the final wash step, 20 μL 3X SDS-PAGE loading buffer was applied to the resin. The magnetic resin allows for separation of solution and resin with a magnetic rack and 15 μL of the supernatant was loaded onto the gel.

2.10 SEC-MALS

SEC-MALS (size-exclusion chromatography - multi-angle light scattering) is a static light scattering technique for determining the absolute molecular mass of a protein (Tarazona and Saiz 2003, Folta-Stogniew, 2006). The technique was employed when the oligomeric state of TINTIN (theoretical MW = 130 kDa) and truncated EA88B (Eaf7 residues 1-260, theoretical MW = 111 kDa) was brought into question and also to assess the polydispersity of the sample. Protein samples were purified as described above and diluted to 0.7-0.8 mg/mL prior to analysis.

2.10.1 Data Collection and Analysis

A DAWN 8+ MALS detector and an Optilab T-rEX refractive-index detector were used to collect data, recording UV_{280}, light scattering (DAWN) and refractive index (T-rEX) signals. These instruments were connected to a HPLC system (Agilent HP 1100), using a Superdex 200 increase 10/300 gel filtration column.

The column was washed with MiliQ water –> 0.5 M acetic acid –> MiliQ –> 0.5 M NaOH –> MiliQ (2 column volumes at each step), before being equilibrated with purification buffer until RI (refractive index) and LS (light scattering) detectors reached a stable baseline. All liquids were degassed within the HPLC system prior to equilibrating the column and detectors.

100 μL of protein sample was injected into the system to initiate the run. The flow rate of the system was set to 0.5 mL/min, with a pressure limit of 30 bar. RI signal and UV_{280} signal was used to verify protein concentration. LS signals were used to determine the absolute molecular weight of species within UV peaks. This data was interpreted using the proprietary ASTRA 6.1 software (Wyatt technology).

2.11 Native Ion Mobility-Mass Spectrometry

The EA88B truncated TINTIN construct (Eaf7 residues 1-260, theoretical MW = 111 kDa) was subject to native mass spectrometry (Native-MS) and ion mobility-mass spectrometry (IM-MS), to gain a better understanding of TINTIN's structure and subunit composition.

This was carried out in collaboration with Charles Eldrid of the Thalassinos lab, who was essential for operation of the Synapt G1 (Waters, UK) and carried out analysis of the raw data that has been interpreted and presented here.

2.11.1 General Theory

Native ion mobility-mass spectrometry (native-MS) is a powerful complimentary technique to high-resolution structural determination techniques (X-ray crystallography, cryo-EM etc.), when investigating the architecture of macromolecular assemblies. Combining conventional mass analysers with ion mobility apparatus, provides information regarding oligomeric state, physical dimensions and flexibility/disorder of protein complexes (Wojnowska et al. 2013, Beveridge et al. 2014, Marklund et al. 2015).

Electrospray ionization (ESI) allows for the "gentle" ionization of proteins injected into mass spectrometry machinery, by releasing protein molecules in a spray of buffer droplets from a highly charged Taylor cone. This is followed by gradual desolvation and accumulation of charge on the gas-phase protein ion. This technique can ensure the overall structure of a protein complex is preserved and is termed ESI-MS or native MS (Fenn et al. 1989, Hall et al. 2012, Beveridge et al. 2014). Native MS is advantageous for projects involving low-yield protein samples, requiring a fraction of the material required for conventional structural analysis techniques (Heck, 2008).

The desolvation process gives rise to a distribution of charges states, whereby protons accumulate on solvent accessible, protonation sites (proton accepting residues). Folded/globular proteins generally have fewer protonation sites exposed to the surface and therefore tend to accumulate fewer protons, compared to extended/disordered proteins, resulting in lower charge states and narrower charge state distributions (CSDs) (Dobo and Kaltashov, 2001).

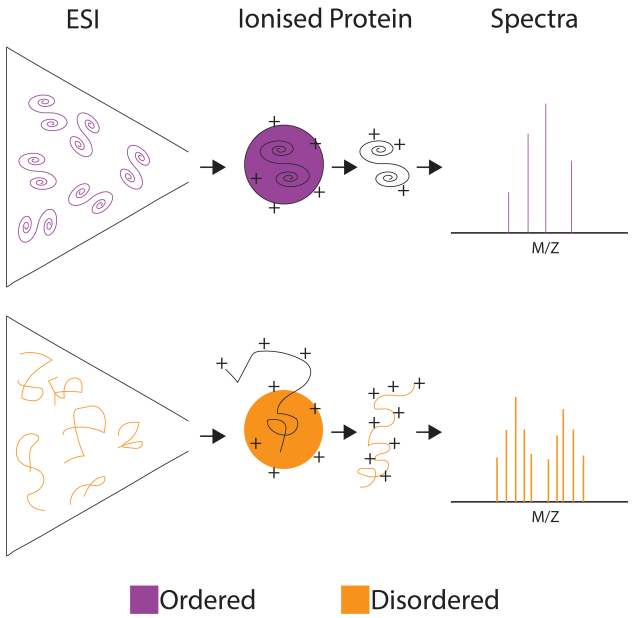

Figure 2.4: Schematic representation on the influence of order/disorder on Native/IM-MS spectra. Solute proteins are ionised via electrospray ionization (ESI), prior to ejection from highly charged ESI tip. Prior to entry into the mass analyser, proteins experience desolvation, retaining some of the charge from the ESI step. Highly ordered proteins have fewer exposed protonation sites compared to disordered proteins, resulting in fewer representative ionized species and a narrower range of collision cross sections.

Ion mobility (IM) adds an additional layer of analysis to mass spectrometry instruments, by providing information on the physical dimensions of the analysed species, in the form of collision cross sections (CCS) (Uetrecht et al. 2010). The relationship between CCS can be approximated as the square of the radius of gyration (R_g), a value typically cited in other biophysical techniques such as SAXS, though this relationship is weakly correlated (Chirot et al. 2010, Beveridge et al. 2014).

The technique works by pulsing ionized proteins (by electric field) through a cell containing inert gas, with larger proteins (with a greater collision cross section) impeded to a greater extent, compared to smaller/more compact proteins. Proteins can produce a distribution of arrival times through the IM-cell, as a consequence of conformational flexibility and variation in charge states (and therefore produce a range of CCS values).

2.11.2 Sample Preparation

TINTIN and protein standards were buffer exchanged into 100 mM and 200 mM ammonium acetate (pH 7.4) solution, respectively. The exchange was carried out using Amicon Ultra centrifugal filtration units (Merck Millipore). Ammonium acetate is a volatile solution, suitable for desolvation during the ESI process, and tends to produce more compact CSDs as a consequence of salt-crowding.

Proteins were diluted to 5-10 μM (TINTIN = 0.7 mg/mL ~6 μM) prior to injection into the machine with gold coated-nESI capillaries.

Protein	MW (kDa)
Adh	146.8
Blac1	70.1
BSA	66.4
Blac2	63.4
Sap10	49.3
Sap5	45.6

Table 2.10: List of proteins used in calibration for IM-MS experiments. The Amphirite software was used to generate a calibration curve (CCS vs arrival time) from these samples, with an R^2 value of 0.98. Arrival times of TINTIN ions were compared to this standard curve to determine CCS values.

2.11.3 Instrument and Operation

A Synapt G1 (Waters, UK) IM-MS device was used throughout the experiments. The IM-cell is positioned between a quadrupole and time-of-flight(TOF) mass analyser. This allows for the selection of ions with specific m/z values for ion mobility experiments, as well as tandem mass spectrometry and collision induced dissociation experiments.

Experimental settings are summarised in table 2.11. Initial native spectra was first inspected to identify the presence of TINTIN species and any TINTIN-associated subunits (individual subunits or dimers).

MS	Capillary (kV)	1.4
	Sampling cone (V)	40
	Extraction cone (V)	1.0
	Backing pressure (mbar)	6.00
	Trap CE (eV)	6
	Transfer CE (eV)	4
	Bias	16
	Mass range (m/z)	1000 - 16,000
Ion Mobility	Source wave velocity (ms^{-1})	300
	Source wave height (V)	0.2
	Trap wave velocity (ms^{-1})	300
	Trap wave height (V)	0.2
	IMS velocity (ms^{-1})	260
	IMS wave height (V)	8.0
	Transfer wave velocity(ms^{-1})	260
	Transfer wave height (V)	8.0

Table 2.11: Summary of general settings during native ion mobility mass spectrometry experiments, using a Synapt G1 (Waters, UK). These were set in collaboration with Charles Eldrid of the Thalassinos lab.

2.11.4 Collision Induced dissociation

Upon identification of a suspected TINTIN ion (+22 charge state), the quadrupole was adjusted to select for this specific m/z species. The selected species was then subjected to collision induced dissociation, by increasing the voltage of trap and transfer components, in order to break the suspected TINTIN complex into its component subunits. Collision voltages increased in increments of 10 V, until the single peak in the spectra split into two or more peaks, indicative of subunit dissociation.

2.11.5 Data Analysis

Mass calculation and peak identification was carried out using MassLynx (v4.1). CCS calculation was determined with Amphirite (Sivalingam et al. 2013) by Charles Eldrid. The relationship between the CCS values and the corresponding MW of proteins, can be described with a linear function (Beveridge et al. 2014).

The linear relationship between CCS and MW for folded proteins is distinct from the linear relationship for disordered proteins. The two functions can be plot and act as standards, to compare experimental data against.

After the median CCS and CSD values of each TINTIN associated species was determined with MassLynx and Amphirite, these values were plot against the linear functions established by Beveridge et al.. Comparison of these values against both functions, informs whether or not the species in question has CCS:MW or Charge state:MW ratios closer to that of ordered or disordered proteins.

2.12 Small-Angle X-ray Scattering

2.12.1 General Theory

For protein samples that cannot be crystallised, or are unsuitable for for structural characterisation by electron microscopy and NMR, SAXS offers a low-resolution alternative that is rapidly produced and can provide insights into protein stoichiometry, general architecture, and physical dimensions (Jacques and Trewhella, 2010).

When a collimated beam of X-rays interacts with molecules in solution, the incident beam may be absorbed, pass through without perturbation, or scattered (Mertens and Svergun, 2010). If there is minimal energy transfer between incident and resultant beams, the X-rays are said to have elastically scattered.
Elastically scattered waves may interact constructively or destructively, producing unique diffraction patterns in the case of protein crystals (Ameh, 2019).

Unlike molecules in a crystal lattice, proteins in solution have no fixed orientation and the resulting scattering signal produced from such experiments, is radially averaged. This contains information regarding the protein's MW and physical dimensions, but loses high-resolution information obtained in X-ray crystallography. These datasets are initially represented as one dimensional curves of scattering intensity I vs. s (the magnitude of the scattering vector/momentum transfer).

The momentum transfer is defined as:

$$s = \frac{4\pi Sin(\Theta)}{\lambda}$$

Where Θ is half the angle between scattered and incident beam, and λ is defined as the wavelength of the incident beam.

For proteins in solution, the radially averaged scattering vector will be the sum contribution from both the solvent and solute. The mean difference in scattering density between solute and solvent is referred to as the "contrast" and is usually low between proteins and their solvent.

Therefore, the contribution to the scattering signal from the sample of interest may be determined, by subtracting a scattering signal from the solvent in the absence of protein.

2.12.2 Sample Preparation and Data Acquisition

Two TINTIN constructs truncated in the Eaf7 C-terminal were analysed by SAXS; EA88A (Eaf7 residues 1-221) and EA88B (Eaf7 residues 1-260).

Data acquisition was carried out with the BM29 BioSAXS facility at ESRF Grenoble, with the aid of Dr Nikos Pinotsis. 60 μL samples were kept at 4 °C in a quartz glass capillary (diameter 1 mm).

All beamlines have a fixed range of s from which data can be collected. The lower limit is particularly important, as it defines the lower limits at which the Guinier analysis (discussed below) is valid. For BM29, the s-range = 0.025 - 5 nm^{-1}.

10 frames were collected for each concentration and each of the buffer blanks (at 1 frame/second). The 10 frames were averaged with the PRIMUS software package (Konarev et al. 2003). The Edna in-house software package compared scattering profiles of each frame and confirmed there were negligible levels of radiation damage.

Constructs were analysed in a concentration series (**Table 2.12**), to determine whether or not the resulting signal was concentration dependent (diluted with purification buffer A). A BSA control (4.5 mg/mL) was also used in the experiment, for MW determination of the species in solution.

Concentration of Samples (mg/mL)		
BSA (control)	EA88A	EA88B
4.5	3.38	3.59
	1.51	1.67
	0.72	0.82

Table 2.12: Summary of concentration series used throughout SAXS experiments investigating the structure of two truncated TINTIN constructs (EA88A and EA88B).

The relationship between scattering intensity and sample concentration, can be defined as:

$$I(0) = N(\Delta\rho V)^2 = \frac{C\Delta\rho^2 v^2 MW}{N_A}$$

Where, $I(0)$ = Forward scattering intensity, N = number of scattering particles per unit volume, $\Delta\rho$ = contrast, C = mass concentration, v = partial specific volume, MW = molecular weight and N_A = Avagadro's number.

This equation shows that low concentrations of high-MW contaminants can disproportionally influence the experimental scattering signal, necessitating high-purity samples for analysis that we evaluated with SEC-MALS.

2.12.3 Analysis of Raw Data

All data analysis was carried out with ATSAS software package v2.8 (Franke et al. 2017). The PRIMUS software suite was used to carry out the following analysis:

The scattering profile ($log(I)$ vs. s) can be used to derive the forward scattered intensity, $I(0)$ and the radius of gyration, R_g. $I(0)$ represents forward scattered intensity, where $2\theta = 0°$. *I(0)* is dependent on the number of particles per unit volume and can be used to calculate the molecular weight of the specimen in question.

R_g, is the radius of gyration, the root-mean-squared distance from an objects centre of mass and gives insights into dimensions of the object. Guinier identified a relationship between both $I(0)$ and R_g, for low values of q (Guinier 1969):

$$I(q) = I(0)e^{\frac{-s^2R_g^2}{3}}$$

Plotting $ln(I)$ vs s^2 therefore produces a line from which both aforementioned parameters can be determined.This relationship is only pertinent between S_{min} (limited by equipment used) and $sR_g < 1.3$ (**Figure 2.5**).

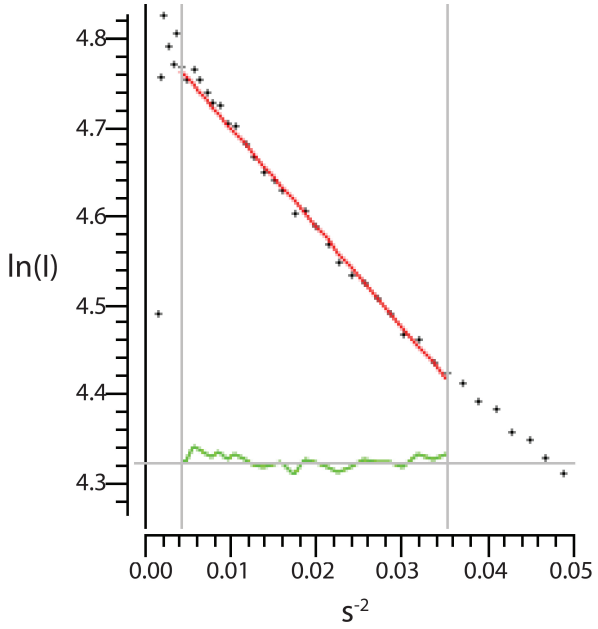

Figure 2.5: Example of background subtracted SAXS data and corresponding Guinier analysis in the PRIMUS software package. Plotting the natural log of intensity from this raw data against the reciprocal of the scattering vector squared, should produce a straight line within the limit of s_{min} and sR_g < 1.3. The R_g and $I(0)$ can be extrapolated from this. Upward curvature in this plot suggests protein aggregation, while a downward curve might suggest repulsive interactions.

$I(0)$ can be used to determine the molecular weight, provided a protein standard of known molecular weight is included in experiments.

The $I(0)$ of the BSA was determined and the MW of TINTIN species was calculated by the following equation:

$$\frac{MW_x}{I(0)_x} = \frac{MW_{standard}}{I(0)_{standard}}$$

Where x is the protein of interest.

Indirect Fourier transform of the scattering profile, with the GNOM software suite (Svergun 1992), produces a *P(r)* (pair-distance distribution) function. This describes the frequency of interatomic distances, within the molecule of interest (Svergun and Koch, 2003)(**Figure 2.6A and B**).

The shape/symmetry of a protein can significantly influence the resulting $P(r)$ function and the plot can provide additional information, not immediately derived from the scattering profile.

For example, the theoretical determination of D_{max} (the maximum dimension of the protein molecule), can be derived from the $P(r)$ function. This is achieved by identifying the r coordinate where $P(r)$ approaches 0 .

$P(r)$ functions can also be used to determine both the R_g and $I(0)$ using the entire dataset. The variance of a $P(r)$ function provides the R_g, while integrating under the curve between $r = 0$ and D_{max}, provides $I(0)$:

$$I(0) = 4\pi \int_{r=0}^{r=D_{max}} P(r)dr$$

Unlike Guinier approximations, the *P(r)* function uses the entire dataset and is considered a more accurate means of determining the physical dimensions of the protein under investigation. However, both methods were used to determine values of R_g, D_{max} and $I(0)$ (and by extension; MW).

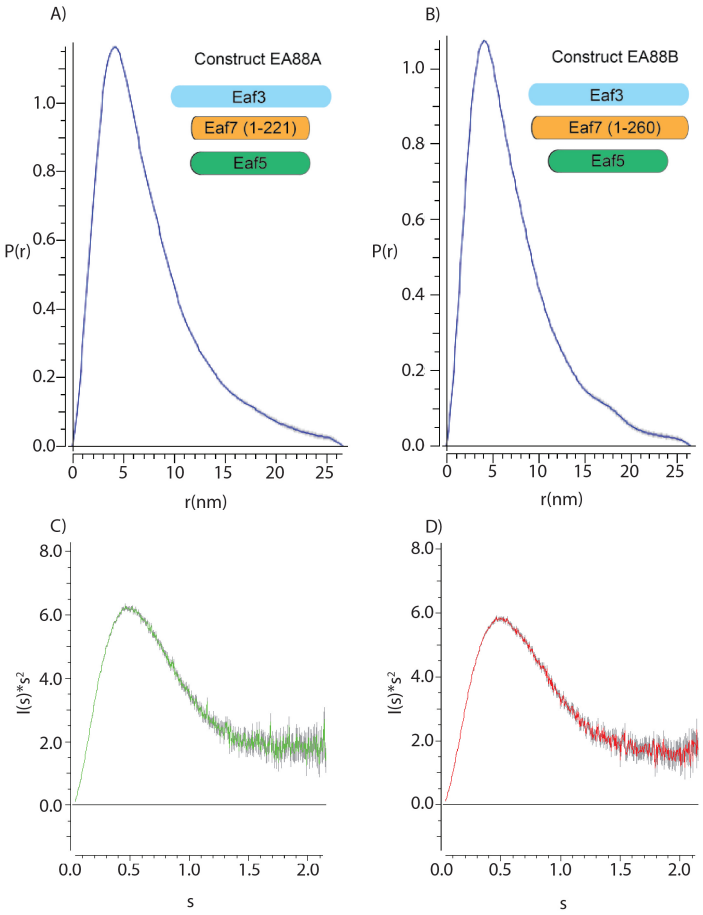

Figure 2.6: A and B) Example $P(r)$ function in the PRIMUS softawre package. The function uses the entire raw dataset (unlike Guinier analysis) and provides values for R_g, D_{max} and $I(0)$. Globular proteins produce normal $P(r)$ distributions, whereas extended conformations produce tailed traces as seen in the figure. **C and D)** Kratky plots of same SAXS data set (Is^2 vs s), which are used to measure the flexibility of protein samples.

The shape of $P(r)$ functions from the data was also considered, as globular proteins are expected to produce normal distributions, whereas extended conformations generally produced asymmetric distributions. Is^2 vs s plots (Kratky plots), were also used to determine whether or not the proteins under investigation showed signs of flexibility. Normal distributions are associated with stable, globular proteins, while asymmetric distributions indicate disordered or partially flexible proteins (Putnam et al. 2007) (**Figure 2.7**).

Integration of the Kratky plot produces the Porod invariant (Q):

$$Q = \int_0^{\infty} I(s)s^2 ds$$

The Porod volume can then be derived, provided the $I(0)$ has been determined, either by Guinier or $P(r)$:

$$V = 2\pi^2 . \frac{I(0)}{Q}$$

For globular proteins, the Porod volume can be approximated as 1.6X the MW of the species in question. Determination of the Porod quotient is sensitive to noise in the high s range, which was observed in the collected data, and *P(r)* function indicated an elongated structure (see Results). As a result, Porod calculated volume is cited in the results chapter, but was not used to determine the MW of the TINTIN complex in solution.

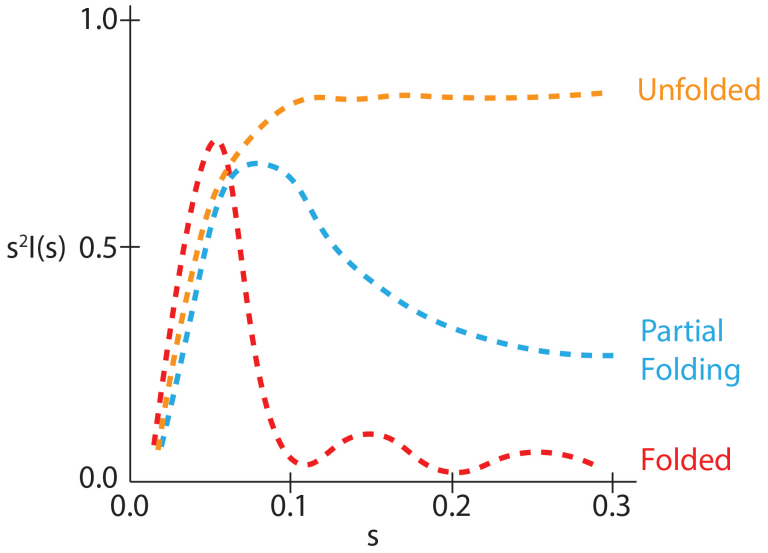

Figure 2.7: Example of Kratky analysis of three proteins, representative of unfolded, partially folded/flexible and folded structures. Integration of the Kratky plot between s = 0 and s = ∞ yields the Porod quotient Q, which can be used to determine the volume of the molecule and MW when combined with *I(0)*.

.

2.12.4 Model Building

.out files produced by GNOM were used to generate *ab initio* models with the dammin software package (Svergun, 1999). dammin works by first creating a search space filled with dummy atoms. The software assigns each dummy atom within the space as either "protein" or "solvent". Dammin then generates theoretical scattering profiles for these models and attempts to match the theoretical curve, with the experimental curve.

The dimensions of the search space were determined by analysis of the *P(r)*, therefore a cylindrical model (due to skewed *P(r)* function) of length 280 Å (determined to be above the predicted D_{max} of ~260 Å) and diameter 60 Å.

Dammin reiterates this process between 10-20 times and generates a log file with χ^2 values that should approach 1. The process was repeated, altering the dimensions of the search space until the statistically best model constraints were applied.

Ab initio models from dammin, are then averaged and filtered with the DAMAVER software suite (Volkov and Svergun, 2003), as follows:

1. **DAMSEL**: Compares all *ab initio* models with an additional program that superimposes structures (SUPCOMB) and statistically assesses the quality of this superposition, by comparing normalized spatial discrepancy (NSD) values*. Outliers are are identified if their NSD score is greater than the mean NSD value + 2 standard deviations.

2. **DAMSUP**: Chooses a reference model based on the best NSD scores (discarding outliers identified by damsel) and generates a list of reference "r" models from these alignments with SUPCOMB.

3. **DAMAVER**: takes all "r" .pdb files and generates a probability map based on the position of atoms within these structures, in the form of a damaver.pdb file.

4. **DAMFILT**: removes low-occupancy atoms from the damaver-generated probability map, to generate a most probable damfilt.pdb model, that satisfies the volume restraints of the damaver.pdb file.

*NSD is a quantitative measure of similarity between 2 .pdb files. Higher similarities result in NSD values that approach 0, whereas NSD > 1 indicates some fundamental difference between the 2 models.

Chapter 3

TINTIN Purification and Crystallography: Results

3.1 Summary

The full-length TINTIN trimer has been bacterially expressed and purified via a Twin-Strep tag on the C-terminal of Eaf7. The final sample appears to be relatively pure when assessed by SDS-PAGE and mass spectrometry analysis validated the presence of the three subunits.

Full-length TINTIN was subjected to crystallisation trials, with the aim to produce a high-resolution structure. The majority of crystal screens show no sign of supersaturation, despite protein concentrations higher than 12 mg/mL. Expression of the full-length complex was inconsistent and limited crystallisation trials at higher concentrations.

The presence of degradation products during purification steps and the inability to crystallize the complex, led to the design of truncated TINTIN constructs. While truncations in the Eaf7 C-terminal improved yield and quality of the sample, additional Eaf5 truncations appear to compromise the integrity of the trimer. Nevertheless, test expression/purification of these constructs, provides insights into the architecture of the complex. The most successful of these constructs were subjected to up-scaled expression and purification and while some promising leads were observed, the complex was still unable to crystallise, despite efforts to further refine expression, purification and buffer conditions.

Finally, pull-down assays were carried out between truncated TINTIN constructs and known (and predicted) TINTIN binding partners. This included peptides mimicking RNA Pol II CTD at various stages of transcription, as well as peptides mimicking H3 N-terminal tail of various methylation patterns. The results of these assays are inconclusive, while low levels of binding are observed (by SDS-PAGE analysis), bands appear faint. However, there may be indications of direct interaction with Pol II CTD.

3.2 Initial Purifications of Full-length TINTIN

After generating the TINTIN-pET28a constructs, Rosetta 2 cells were transformed and incubated in the following expression conditions, to survey for optimal TINTIN expression (**Table 3.1**)

Condition	Pre-Induction	Induction OD	[IPTG] (mM)	Post-Induction
AI1	37 °C	0.5	-	18 °C(24 hours)
AI2	37 °C	0.5	-	37 °C(16 hours)
LB1	37 °C	0.5	0.2	37 °C(4 hours)
LB2	37 °C	0.5	0.2	18 °C(18 hours)

Table 3.1: Summary of test expression schemes used in the production of TINTIN. All cells were shaking at 200 rpm throughout their expressions. Schemes in the table are displayed as: pre-induction temperature, OD cells reach before induction, IPTG induction concentration, post-induction conditions cells are kept at. Conditions listed as AI1 and AI2 used auto-induction media and therefore did not require IPTG driven induction.

Small-scale (50 mL) expressions were harvested, flash frozen with liquid nitrogen for storage, then subjected to Strep-Tactin purification. SDS-PAGE analysis of these purifications show that all three TINTIN subunits are expressed, irrespective of the conditions cells are incubated in, with each lane producing a similar pattern of bands (**Figure 3.1**).

It was also noted that despite the Strep-tag's high-affinity and specificity, the elution lanes produce many contaminant bands. Condition AI2 yields the best result in terms of intensity of suspected TINTIN subunits, though it also appears to have the highest concentration of contaminant proteins. The Eaf7 subunit runs ~20 kDa higher in SDS-PAGE analysis (just above 70 kDa), with respect to other proteins of equivalent MW. Eaf3 (45 kDa) shows a more conventional migration pattern, running between 40 kDa and 55k DA, and Eaf5 runs close to 35 kDa.

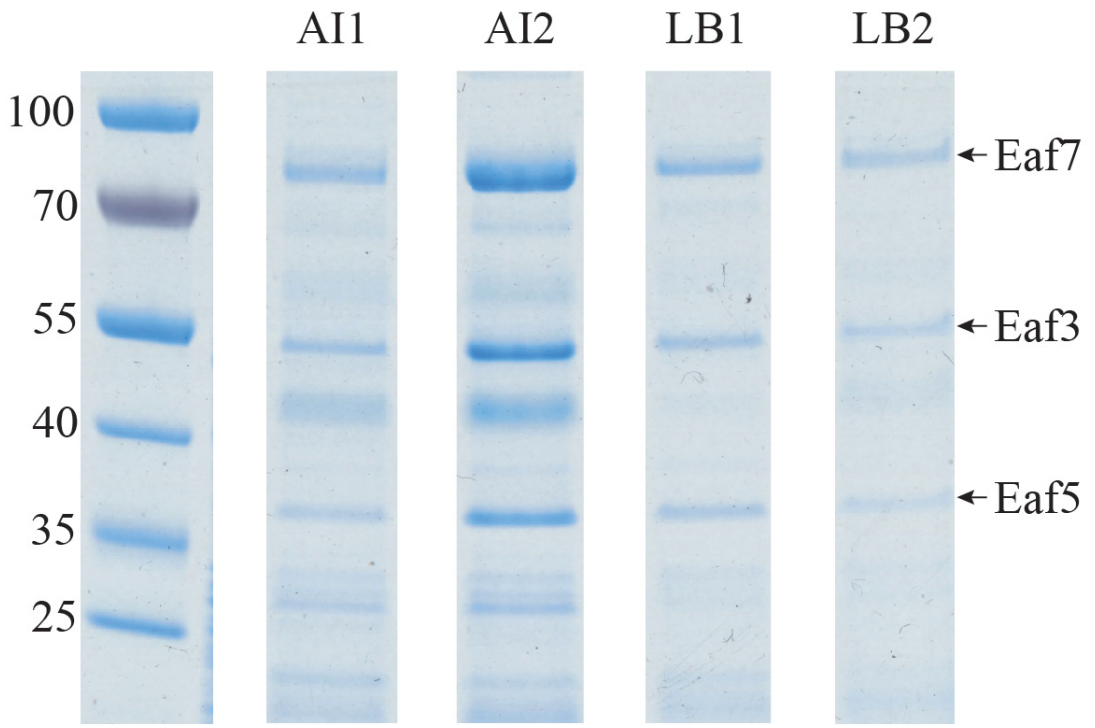

Figure 3.1: SDS-PAGE analysis of small-scale (50 mL) expressions/purifications of TINTIN-pET-28a in Rosetta 2 cells, comparing LB and auto-induction(AI) media based expressions. Lanes are grouped according to the expression scheme they represent (see Methods) and correspond to the conditions listed in Table 3.1. Lanes contain 50 μL loads, from a mixture of 10μL 6X SDS-buffer + 50 μL Strep-TactinXT beads. Condition AI2 appears to produce the highest concentrations of TINTIN subunits (based on band intensity). All lanes show high levels of contamination, despite the high-specificity of the purification system employed.

3.3 Strep-Tactin Purification Optimization

Test expression results show that the optimal conditions for TINTIN expression, was to incubate cells in AI media at 37 °C (200 rpm) for 24 hours. This was up-scaled to 12 L expressions, producing cell pellets with mass in the range of 150 g.

Early purifications from these pellets were limited by inefficient lysis by sonication (60% amplitude, 2 s/2 s (on/off) for 10 minutes), and a sub-optimal Strep-purification step, using 1.5 mL of resin and eluting with 5 x 2 mL aliquots of elution buffer (**Figure 3.2A**). Samples taken from elutions of the Strep-Tactin purification step, produced faint bands corresponding to suspected TINTIN subunits. Furthermore, elutions from this step were highly contaminated by the products of TINTIN degradation, which were verified by mass spectrometry.

Later purification attempts (**Figure 3.2B**) involved a series of adjustments to improve yield and sample quality. These adjustments included; doubling the concentration of protease inhibitor cocktail (to reduced degradation), using 3 mL Strep-TactinXT resin which is designed specifically to bind twin-Strep tags with higher affinity and capacity.

These adjustments dramatically increased the concentration of TINTIN subunits purified after Strep-TactinXT purifications (**Figure 3.2B)**. Eaf5, Eaf7 and Eaf3 bands were equally intense, and the final quantity of purified TINTIN (after IEX and SEC purification steps) was ~1.8 mg of protein.

Despite these improvements, two outstanding issues became apparent. Firstly, the Strep-TactinXT purification step produced relatively impure sample, as seen in Figure 3.2B. Secondly, while the intensity of TINTIN bands increased, we found that even after the addition of up to 14 mL of elution buffer (containing 50 mM biotin), elution was incomplete. This indicated that a large percentage of the expressed complex was still bound to the resin and that the elution step required further optimisation.

This was supported by the fact that regenerated resins (washing the resin with a low concentration sodium hydroxide solution), showed reduced binding of TINTIN in future purification attempts.

Further optimisation of the Strep-purification step addressed the suspected degradation of the purified sample and maximised the elution of TINTIN from the Srep-TactinXT resin (**Figure 3.2C**). To reduce the degradation issue, we included protease inhibitor in all buffers (including the elution buffer) at 2X the manual recommended concentration (1 tablet per 100 mL of buffer).

To ensure that all of the TINTIN sample was cleared from the resin, we resuspended the resin in 20 mL of elution buffer, and left it rotating at 4 °C for 1 hour, before eluting all 20 mL in one step. An additional 20 mL of buffer was added and the resuspension/incubation was repeated. The pooled 40 mL of buffer was found to contain up to 27 mg of protein (measured by Bradford assay), a tenfold increase in yield from initial Strep-Tactin purification attempts (**Figure 3.2C**).

We verified that the resin had been stripped of TINTIN, by adding an additional 5 x 2 mL aliquots of elution buffer, finally observing a tapering of band intensity with consecutive applications (**Figure 3.2C**). To confirm complete elution, the resin was incubated overnight in elution buffer and no protein was detected in SDS-PAGE analysis (data not included).

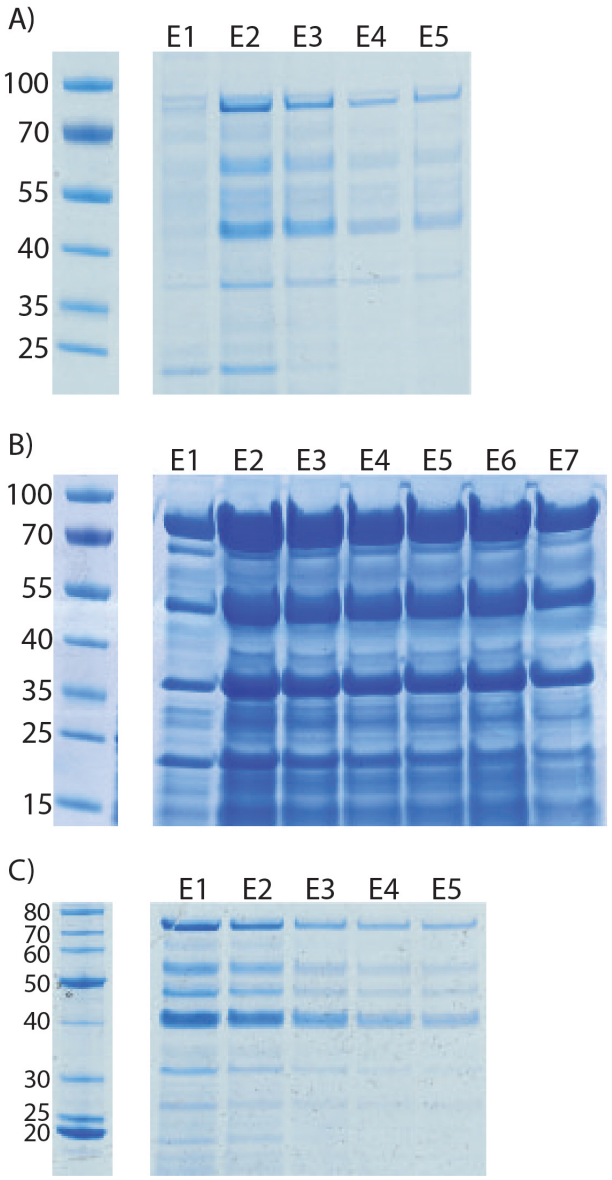

Figure 3.2: SDS-PAGE analysis showing the iterative optimization of Strep-Tactin purification stage. Lanes correspond to consecutive elution fractions (30 μL loads, from a mixture of 30 μL of eluted protein and 6 μL loading dye). **A)** Representative of early elution attempts while all TINTIN subunits. The faint bands reflect low-yields as a result of suspected degradation of the trimer. **B)** Representative of purifications after use of Strep-TactinXT resin. There is a large abundance of contaminating bands in every lane, but TINTIN subunits are clearly distinguishable. Despite the addition of 7 x 2 mL of elution buffer, the intensity of these bands does not appear to taper, suggesting that a large proportion of TINTIN remained bound to the resin. **C)** Representative of optimised elution scheme, using XT resin. Lanes E1 to E5 represent addition of 2 mL fractions of elution buffer added consecutively, after incubating the resin with 2 x 20 mL of elution buffer for 2 hours (Data not included).

3.4 Further Purification

The theoretical pI of TINTIN did not reflect the proteins behaviour during ion exchange chromatography. With the pI of sample predicted to be ~5 and the buffer of our purification remaining at pH 8 throughout all steps, it was expected that the sample would be anionic and bind the Mono Q (5/50 GL) 1 mL column.

However, TINTIN purified from the Strep-purification step, does not appear to bind the resin and a strong UV_{280} signal is detected in the flow-through after sample application. No protein eluted after application of a high-salt buffer B. It appears that other contaminants that are present in the initial sample, elute upon application of the high-salt buffer gradient (data not included). Due to this result, the flow-through from the Mono Q 5/50 GL (which should have contained TINTIN) was loaded onto a Mono S 5/50 GL (**Figure 3.3A and B**), a cation-exchange column.

Unlike the Mono Q column, the flow-through of the Mono S appears to be completely clear of protein, while we detect a strong UV_{280} signal during the high-salt elution gradient (~16% high-salt buffer B). This peak overlaps with 2 smaller peaks, which are always present, despite optimisation of earlier stages of the purification (**Figure 3.3A**).

The major peak was confirmed by SDS-PAGE analysis to contain TINTIN subunits, albeit with less contaminating bands (**Figure 3.3B**). Many of the contaminating proteins seen in the Strep-purification step, are separated into the smaller peaks and the height of these peaks correlated with the extent of TINTIN degradation.

As seen in Figure 3.3, while the major peak does contain TINTIN, there is certainly overlap with the subsequent peaks. Eaf7 bands appear to reduce in intensity with each fraction, while contaminating bands increase, supporting the notion that TINTIN subunits (more specifically Eaf7) were being degraded.

While there is certainly full-length TINTIN subunit found in these peaks, a compromise between sample yield and purity had to be made for the sake of crystallography trials and biophysical characterisation. As a result, the first five 0.5 mL fractions in the major 280 nm peak were pooled and further purified by SEC.

Initially, due to the low-yield of protein after IEX, the Superdex 200 Increase 10/300 GL (1CV = 24 mL) was used, but was increased to a Hi Load Superdex 200 16/600 (1 CV = 120 mL). TINTIN elutes in the first major 280 nm peak from these runs (**Figure 3.3C and D**), followed usually by two additional peaks which contain contaminating proteins.

Again, the height of these contaminating peaks was usually a reflection of the quality and yield of the final product. The higher the contaminant peaks, the greater levels of degradation that were observed to have occurred by SDS-PAGE analysis.

The first peak is sufficiently pure for pooling and concentrating for crystallisation trials. However, concentration of the sample usually resulted in loss of 30-50% during spin-concentration steps. This issue was never resolved, despite several attempts at optimising the membrane of the spin-concentrators, MW cut-off and even the addition of stabilising agents such as glycerol.

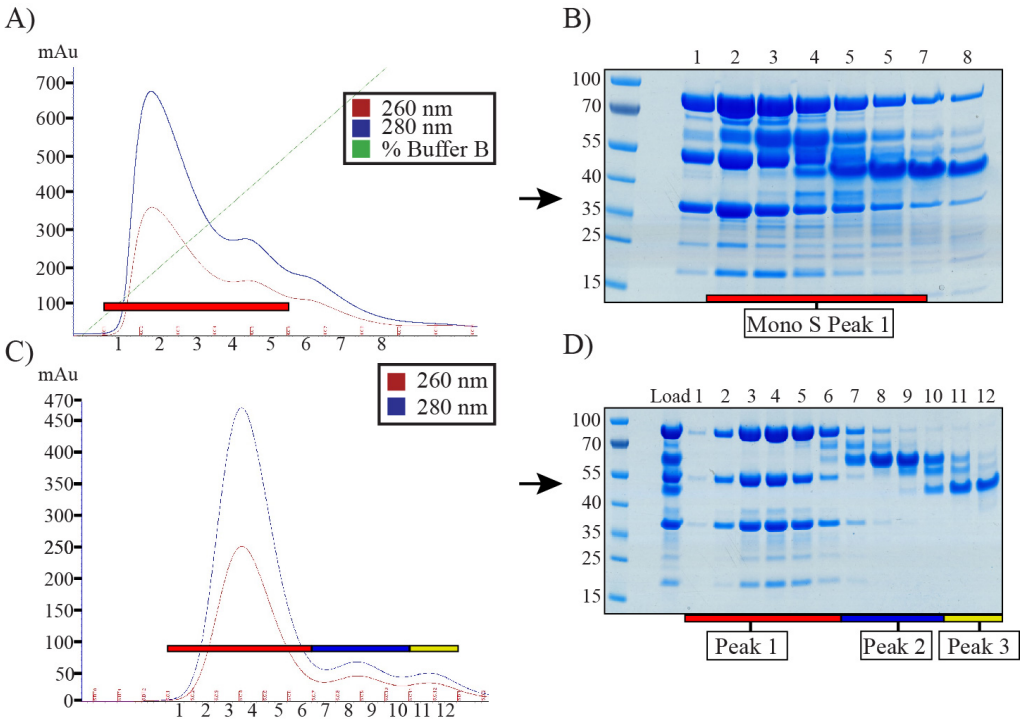

Figure 3.3: Chromatograms of the IEX and SEC steps of TINTIN purification and their corresponding SDS-PAGE gels. Each curve has been colour coded and bars on the bottom of each trace match the lanes labelled in the adjacent gels. **A)** Chromatogram of the TINTIN sample loaded onto a Mono S 5/50 GL, after being cleaned by a Mono Q 5/ 50 GL. The first major peak elutes during the application of the high-salt (2 M NaCl) gradient (16% buffer B). Fractions that correspond to this peak were sampled for SDS-PAGE analysis (highlighted by a red bar). **B)** SDS-PAGE analysis of the IEX chromatography step, using a Mono S 5/50 GL. Lanes above the red bar correspond to peak 1 in **A** and were pooled for further purification. **C)** Chromatogram of the final SEC step of the purification, using a Hi Load 16/600 Superdex 200 pg. Bars beneath the trace correspond to the bars in D. **D)** SDS-PAGE analysis of SEC purification step. The TINTIN trimer is clearly visible in the first peak with some minor contaminants. The first peak is consistently produced during these purifications and was pooled for concentration and crystallisation trials.

3.5 Validation of TINTIN Purification

To confirm that the bands being observed were TINTIN subunits, test purifications were carried out at a smaller scale. Suspected TINTIN bands in the final SDS-PAGE analysis were excised, along with two intense bands that were consistently appearing in this final stage of purifications (**Figure 3.4**).

The bands were sent to the St. Andrews mass spectrometry unit for protein identification, verifying the presence of TINTIN subunits. It was also found that the two intense bands that elute after the complex, are degradation products of the Eaf7 subunit.

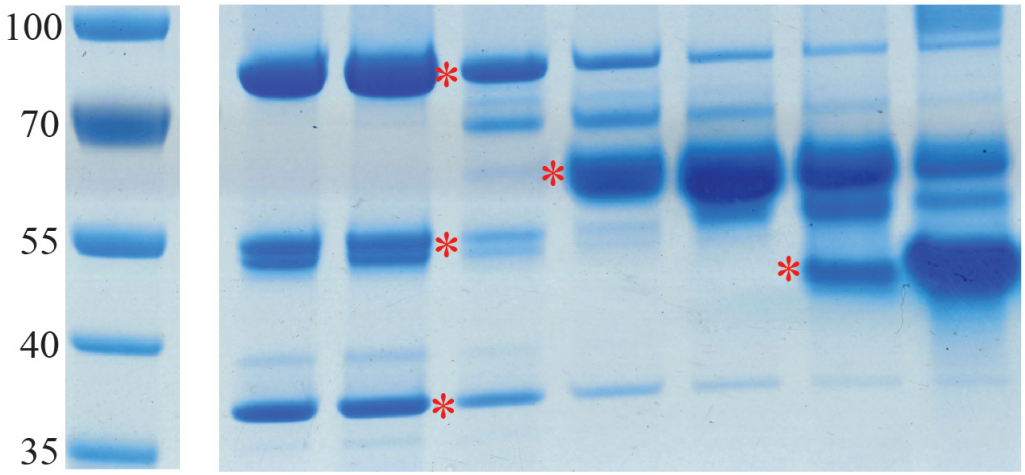

Figure 3.4: SDS- PAGE analysis of SEC purification from TINTIN purification, marked bands were sent to St. Andrews mass spectrometry service for protein identification. The band between 100 kDa and 70 kDa is identified as Eaf7, the two bands close to 55 kDa were identified as Eaf3, the band above 30 kDa was identified as Eaf5. The two additional contaminant bands, were shown to be C-terminal degradation products of Eaf7.

3.6 Expression/Purifications Optimisation

Despite the relatively high-yields of the TINTIN complex that were being expressed in AI media, it was noted that there was an abundance of degradation products observed in the initial stages of the purifications, particularly of the Eaf7 complex. At this stage, crystallisation trials were yielding no successful leads and it was proposed that the complex was too dilute, or that the contaminants seen after the final gel filtration step, were somehow impeding the crystallisation process.

Moreover, expression of the full-length complex was extremely inconsistent, with discrepancies in protein yield between identical expressions/purifications, despite no obvious change in growth rate in Rosetta 2 cells. The integrity of the expression system and construct were assessed by test expression of other constructs and sequencing (data not included), but the source of this inconsistency was never identified with certainty.

With the aim to improve the consistency of yield and purity, LB and Terrific broth media were tested for their efficacy, in expressing the TINTIN complex. In both instances, 6 x 1 L LB/TB (in 2 L flasks) was inoculated with 2 mL of pre-culture grown overnight (flasks also contained kanamycin and chloramphenicol as described in the Methods). Flasks were incubated until OD = 0.6-0.7 before induction with 0.2 mM IPTG, and were further incubated at 18 °C for 18 hours.

Pellets produced from either type of expression, were subjected to the same optimised purification protocol, used for AI media expressed TINTIN. The complex behaves identically throughout the purification and the final result observed suggests no observable improvement over AI media expression (**Figure 3.5**). Moreover, we find that the yield of purified TINTIN, is dramatically lower than the original expression and purification scheme, for equivalent volumes of media used.

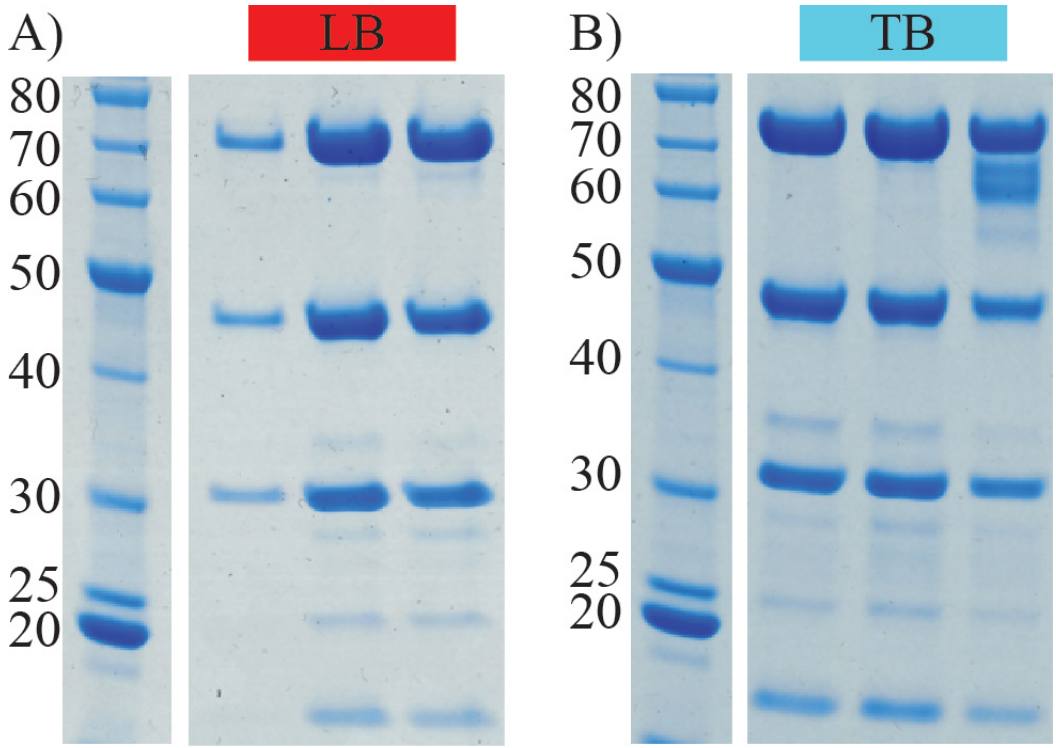

Figure 3.5: SDS-PAGE analysis of TINTIN purifications, where the complex has been expressed in LB or TB. In both instances, TINTIN subunits are clearly expressed and purified, however contaminant bands are still observed. The yield of these purifications was also significantly lower than those produced by AI media expressions. **A)** LB expressed/purified TINTIN. 50 μL loads from SEC fractions (from a mixture of 50 μL sample + 10 μL 6X dye). **B)** TB expressed/purified TINTIN. 50 μL samples from the SEC elutions were taken and mixed with 10 μL loading dye. 50 μL of this was loaded onto the gel.

3.7 Full-length TINTIN Crystallisation Trials

Crystallisation screens were carried out in parallel with the TINTIN purification optimisation (**Table 3.2.**). The concentration of TINTIN for each round of crystallisation attempts, reflected the maximum yield of the trimer achievable at that point in time. As a result, increasing concentrations of TINTIN were used with each attempt.

Commercially available, sparse matrix screens were used to cover a wide range of crystallisation conditions. TINTIN samples were mixed with each condition (150 nL sample + 150 nL buffer) by Mosquito crystal into MRC 2-well crystallisation plates. These were stored at 16 °C in a Rigaku minstrel DT imager and gallery DT plate hotel. Each plate was immediately imaged upon loading onto the hotel platform, allowing for detection of any immediate crystal formation. Samples were then periodically imaged by the device and saved on a database for review. Early purification attempts were limited in their yield and purity, but gradual improvements to purification protocols steadily improved this.

Sparse Matrix Screen	Concentration (mg/mL)			
	4.28	7	12.4	10 (4 °C)
JCSG-plus	X	X	X	X
Pact	X	X	-	X
SaltRX HT	X	-	X	X
Index	X	X	X	X
Structure Screen 1	X	X	X	X
Midas	X	-	-	-
ProPlex	X	X	X	-
Morpheus® HT-96	-	X	X	X

Table 3.2: Summary of crystallisation attempts. Each commercially available sparse matrix screen is given an 'X' if it was used for a corresponding TINTIN preparation. Protein sample and buffers from each screen were mixed 1:1 (v/v) by Mosquito robot for accuracy and reproducibility, the resulting plates were stored at 16 °C or 4 °C and regularly imaged to search for the presence of crystal formation.

3.7.1 *Index*

The vast majority screens containing 4.28 mg/mL TINTIN appear completely clear, even after 240 days of incubation, with a minority (~15%) showing signs of supersaturation (in the form of precipitation, phase separation etc.). We speculated that this was due to the aforementioned limitations of the TINTIN purification process, which was further optimised to 7 mg/mL, 10 mg/mL and then 12.4 mg/mL (at a volume sufficient to prepare > 5 sparse matrix screens).

Even after improvements in yield, crystals were not observed to form after 240 days incubation, with most wells clear (~80% on average across all screens). Six screens were incubated at 4 °C, using one of the highest concentrations of TINTIN produced (10 mg/mL) (these plates required manual analysis, without the aid of a UV detector).

While the majority of wells in the 4 °C plates remained clear, there was a marginal increase in the number of wells containing some form of precipitation, or phase separation. Once again, crystal formation was not observed across any of the 6 sparse matrix screens. After 3072 tested conditions, covering a spectrum of buffer conditions, temperatures and increasing sample concentrations, no indication of crystal formation was ever obtained.

Despite this, several distinct wells in the Index screen were scrutinised for their buffer composition, after certain buffer components were shown to repeatedly coincide with sample precipitation. Figures 3.6A and B show two wells produced from 12.4 mg/mL TINTIN, incubated at 16 °C after 17 days of incubation.

The figure shows small, round structures that UV detectors suggest contain protein. The drop in Figure 3.6B looks particularly interesting, due to the hard-edged appearance of these aggregates.

Earlier attempts with this particular screen (using lower concentrations of TINTIN) do not show any indication of precipitation in these particular wells. It was noted that both wells contain polyethylene glycol (PEG) 3,350, which immediately led to review of any other well within the screen, that contained the compound.

Of the 39 wells containing varying percentages of PEG 3,350, only one other well (containing 0.1 M succinic acid, 15% PEG 3,350 with an average pH of 7.3) shows a similar level of precipitation and morphology.

Comparison of these three wells, revealed that they each contained buffering components with similar functional groups. Sodium malonate, sodium citrate tribasic dihydrate and succinic acid, are derivatives of Krebs cycle compounds. Examining the screen once again, revealed that wells that contained similar compounds, were more likely to contain precipitated protein. There were no other trends in pH, salt concentration or precipitant observed (**Table 3.3.**).

Interestingly, this trend of citric acid cycle derivatives driving TINTIN precipitation, is observed in other sparse matrix screens (**Figure 3.6C and D**). For example, in a well from the ProPlex crystal screen, containing 0.05 M calcium acetate, 0.1 M sodium cacodylate, pH 6, 25% (v/v) MPD (**Figure 3.6C**).

Another well in the Structure screen, containing: 0.1 M sodium citrate, pH 5.6, 20% (w/v) polyethylene glycol 4000, 20% (v/v) 2-propanol, produced similar results observed in the Index screen (**Figure 3.6D**).

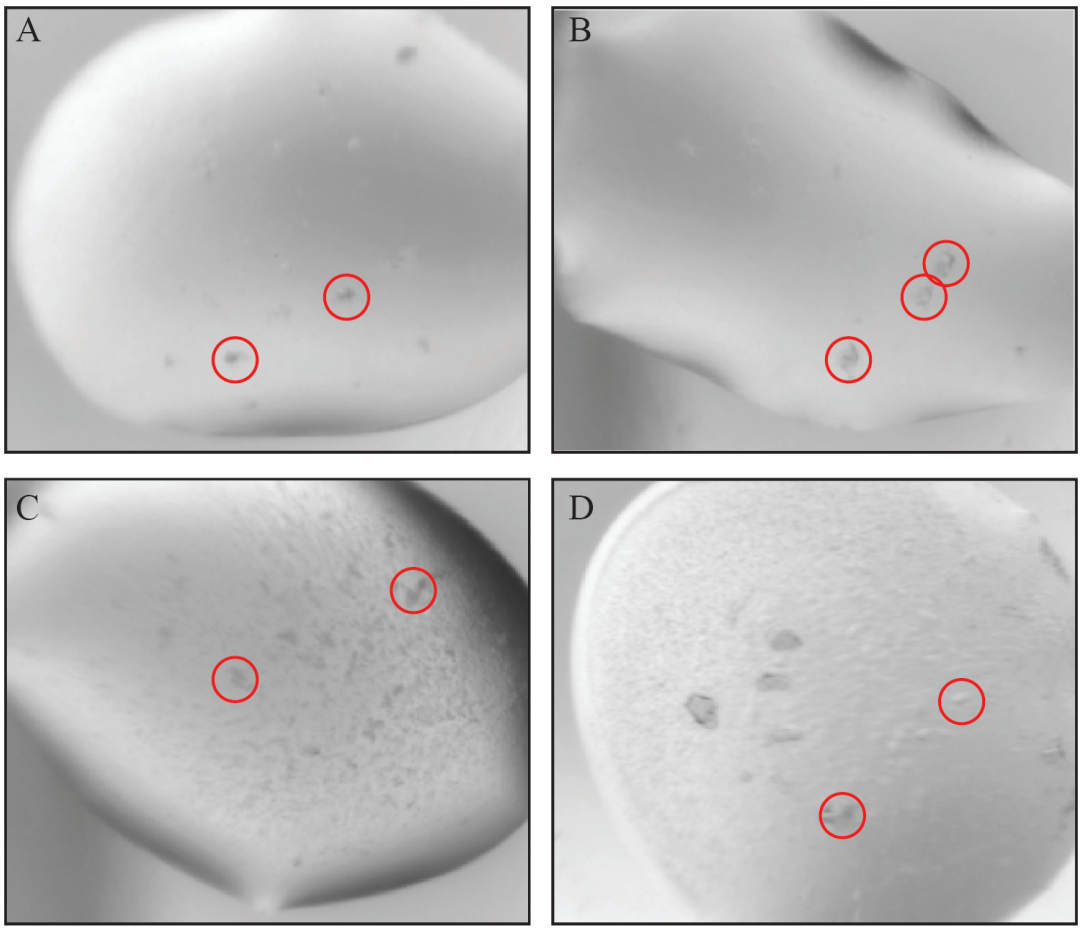

Figure 3.6: Wells containing 150 nL 12.4 mg/mL purified TINTIN and 150 nL of buffer from the Index sparse matrix screen, after 17 days incubation at 16 °C. **A** and **B** are from the Index screen, while **C** and **D** are from the ProPlex and Structure screens, respectively. Protein aggregates have been circled in red. **A)** This well contains 0.2 M sodium malonate pH 7, 20% (w/v) polyethylene glycol 3,350, average pH = 7.4. **B)** 0.2 M sodium citrate tribasic dihydrate, 20% (w/v) polyethylene glycol 3,350, average pH = 8.3. The structures produced in this well appear hard edged and were initially mistaken for protein crystals. **C)** This well contains 0.05 M calcium acetate, 0.1 M sodium cacodylate, pH 6, 25% (v/v) MPD. **D)** Contains: 0.1 M sodium citrate, pH 5.6, 20% (w/v) polyethylene glycol 4000, 20% (v//v) 2-propanol. After 240 days incubation at 16 °C, no further changes in morphology were observed for either drop.

Compound	Concentration (M)	pH	Result	PEG 3,350
Citric acid (+2M Ammonium Sulfate)	0.1	3.5	High Precipitate	-
Citric acid (+3M NaCl)	0.1	3.5	High Precipitate	-
Sodium Citrate Tribasic Dihydrate	1.4	7.5	Precipitate	-
Sodium Citrate Tribasic Dihydrate	0.2	8.3	Precipitate	20%
Ammonium Citrate Tribasic	1.8	7	Low Precipitate	-
Ammonium Citrate Tribasic	0.2	7	Clear	20%
Succinic acid	0.1	7	Precipitate	15%
Succinic acid	0.8	7	Clear	-
Succinic acid	1	7	Clear	-
Sodium Malonate	2.4	7	High Precipitate	-
Sodium Malonate	1.1	7	Clear	-
Sodium Malonate	0.2	7	Precipitate	20%

Table 3.3: Summary of analysis of Index sparse matrix screens, containing derivatives of the citric acid cycle. Concentrations of each compound are described, as well as the pH and PEG 3,350 content. PEG 3,350 and pH levels do not appear to be the determining factor in the extent of precipitation. 75% of the conditions described here show some level of saturation of the TINTIN sample. We verified that the precipitation observed contained protein by UV detector within the Rigaku storage device.

3.7.2 *Morpheus*

One other sparse matrix screen produced a promising trend of results. The Morpheus HT- 96 screen produced results similar to those seen in Figure 3.6, but did not contain citric acid cycle derivatives. As with the Index sparse matrix screen, when TINTIN was applied to this screen at concentrations < 12.4 mg/mL. These wells appear completely clear, indicating that the phenomenon observed is dependent on protein concentration.

The well in Figure 3.7A contains 0.12 M ethylene glycol, 0.1 M buffer system 1, pH 6.5 50% (v/v) precipitant mix 1. The well in Figure 3.7B contains 0.12 M monosaccharides, 0.1 M buffer system 2, pH 7.5, 50% (v/v) precipitant mix 1. The common compounds in both wells are found in 'precipitant mix 1' which consists of: 40% (v/v) PEG 500 MME and 20% (w/v) PEG 20,000. As a result, wells that contained this precipitant mix were analysed to search for signs of precipitation/supersaturation.

Of the 24 wells that contained the precipitant mix 1, none show similar results to the two described in **Figures 3.7A and B.** However, in 20 drops containing these precipitant compounds, phase separation is observed (**Figure 3.7C**). This occurs at T_0 without any change over the course of 240 days. No other trend in pH levels, precipitant mixtures, buffering compounds, or salt concentration was observed in the Morpheus screen (or any other sparse matrix screens).

Some unique conditions produced similar results to the drops displayed in Figures 3.6 and 3.7. The chemical composition of these drops did not appear to belong to any identified trend, but nonetheless, show the protein sample reaching supersaturation (using 12.4 mg/mL TINTIN in the preparation of these drops) (**Figure 3.7D**).

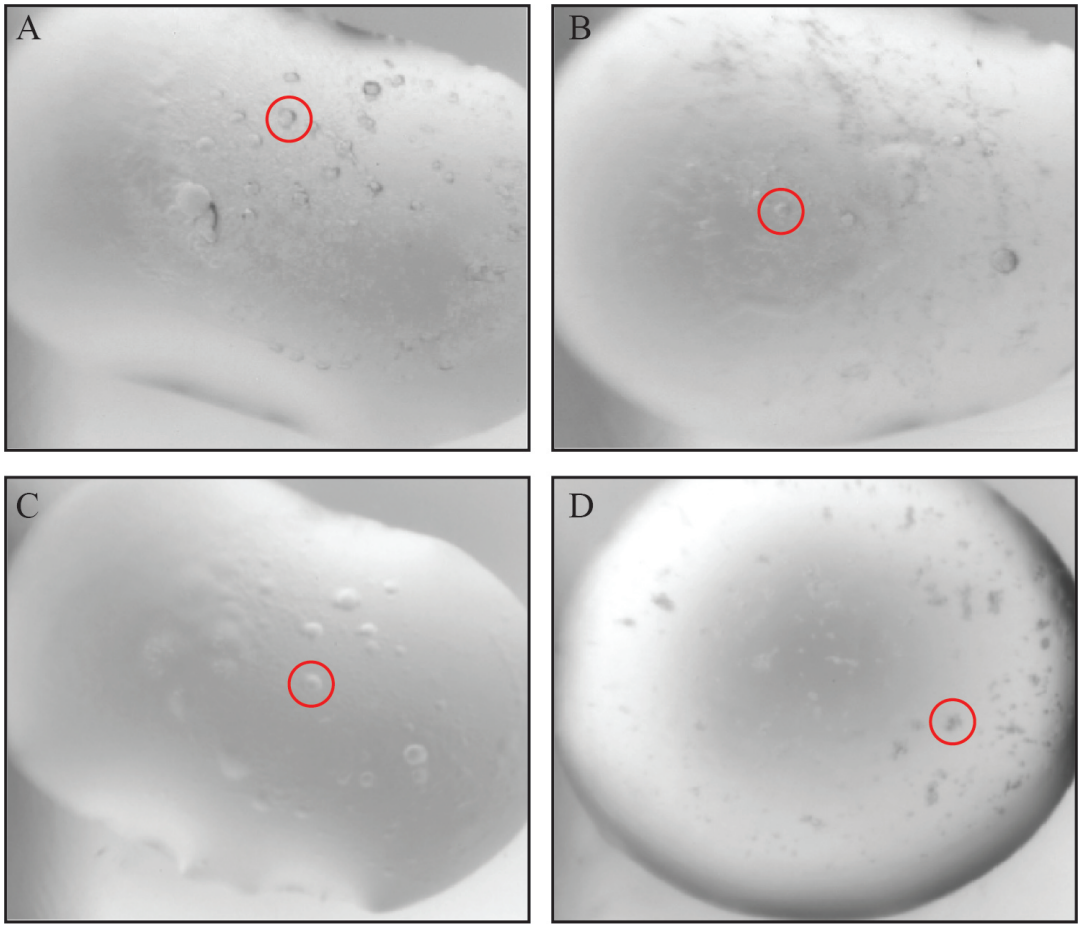

Figure 3.7: Drops produced from the Morpheus® HT-96 screen, containing 150 nL buffer + 150 nL 12.4 mg/mL TINTIN incubated for 17 days at 16 °C. The round structures (circled red) were confirmed to contain protein by UV detector. **A)** Buffer in this well contains: 0.12 M ethylene glycol, 0.1 M imidazole, 0.1 M MES, pH 6.5, 20% (v/v) PEG 500 MME and 10% (w/v) PEG 20,000. TINTIN appears to form round clusters, some of which appear to possess hard edges. **B)** Buffer in this well contains 0.12 M monosaccharides, 0.1 M HEPES, 0.1M MOPS, pH 7.5, 20% (v/v) PEG 500 MME and 10% (w/v) PEG 20,000. **C)** This is representative of 20 other drops from the Morpheus screen, containing the same precipitant mixture. The buffer contains: 0.12 M ethylene glycol, 0.1 M HEPES, 0.1 M MOPS (pH 7.5), 20% (v/v) PEG 500 MME and 10% (w/v) PEG 20,000. These drops also form immediately, and show no morphological change over a 240 day period. **D)** This drop is from the ProPlex screen and contains: 0.05 M magnesium chloride, 0.1 M MES, pH 6.5, 10% (v/v) 2-propanol and 5% (w/v) PEG 4000. This is representative of a small subset of results, across all sparse matrix screens used, with no apparent trend between buffer composition and protein precipitation.

3.8 Limited Proteolysis and Design of Truncated

The lack of success in crystallising the full-length TINTIN complex necessitated a change in approach. High levels of degradation were also observed during the purification process, irrespective of the expression scheme used and optimization attempts. Limited proteolysis was carried out to identify a truncated, but stable TINTIN core and also to identify key components for TINTIN assembly.

Purified TINTIN was mixed with 4 proteases at varying ratios and lengths of time at room temperature. Reactions were quenched with SDS-PAGE loading dye, and the results analysed by electrophoresis (**Figure 3.8A**). Both α-chymotrypsin and trypsin show progressive levels of degradation of the trimer as incubation time increases, while papain completely destroys the sample and subtilisin consistently produces a smear of bands.

Distinct bands that were consistently produced throughout optimization attempts, were excised and sent for analysis to the St. Andrews mass spectrometry service. This analysis included determination of the boundaries of each peptide detected, allowing for the design of truncated constructs to be be expressed and purified for crystallography (**Figure 3.8B**).

Eaf7 degradation from the C-terminal was observed and degradation from the N-terminal of Eaf5 was also detected. Eaf3 species showed minor degradation from both termini, but the subunit remained generally intact. Considering these results, in combination with disorder prediction methods (**Appendix A**) that indicated that ~70% of Eaf7 is disordered from the C-terminal, a series of truncated constructs were designed (Methods). Each of these constructs contained a C-terminal Twin-Strep tag, identical to that of the full-length construct. In parallel to this, the chromodomain domain of Eaf3 was also designed with C-terminal Twin-Strep tags, for the purpose of TINTIN pull-down assays.

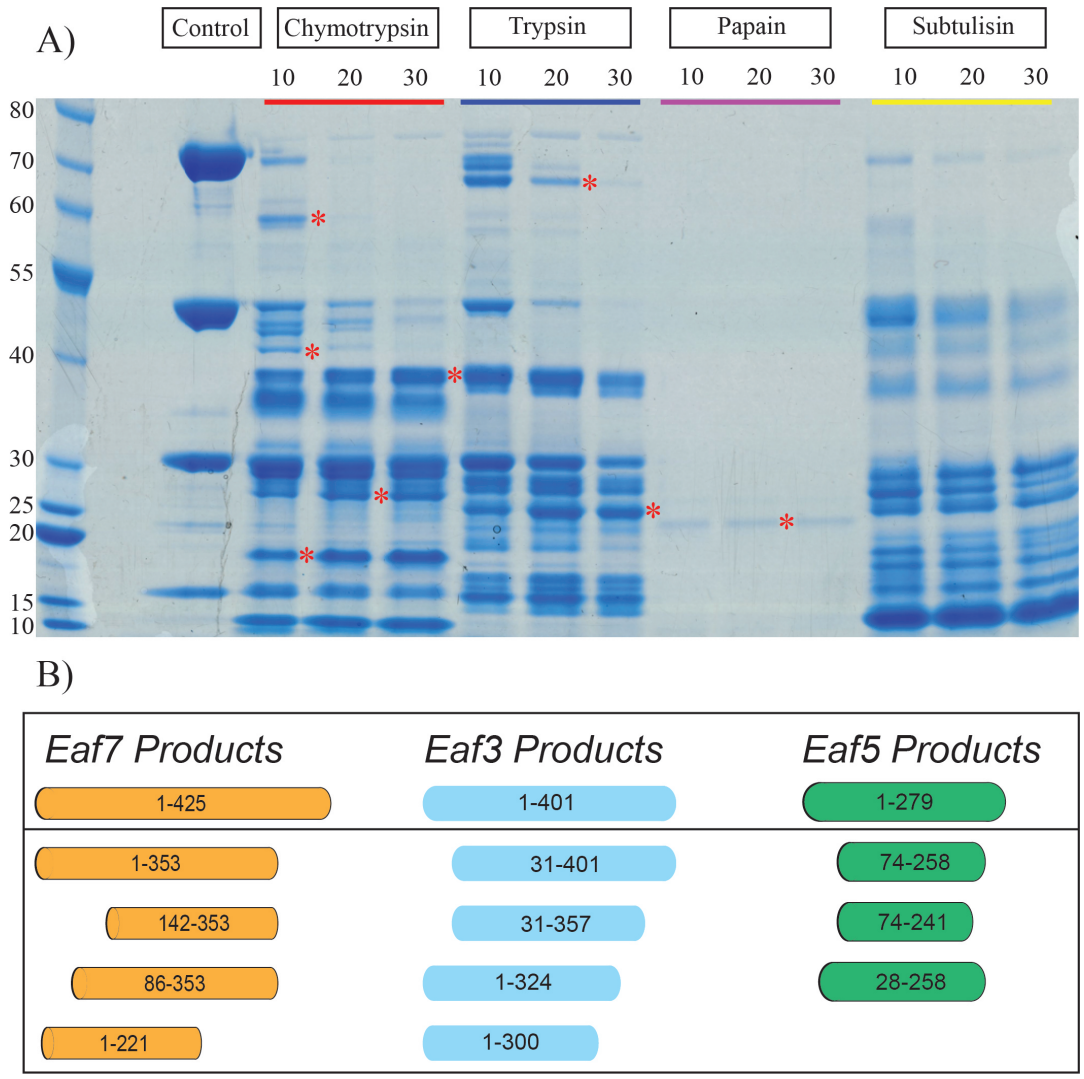

Figure 3.8: A) SDS-PAGE analysis of optimised limited proteolysis experiments. Values above each lane correspond to incubation time (minutes), between TINTIN and protease at room temperature. In the α-chymotrypsin and papain reactions, 140 ng of protease was present, whereas trypsin and subtilisin reactions contained 2.8 ng of protease. Reactions were quenched with 8 μL of 6X SDS loading dye, at varying time points. 20 μL of these mixtures were loaded onto the gel. α-chymotrypsin and trypsin lanes show the best results, with progressively increasing levels of degradation with incubation time. Bands marked with red, were excised and sent to St. Andrews for identification by mass spectrometry. **B)** The boundaries identified by the mass spectrometry analysis, are listed beneath the full-length subunits in a schematic format.

3.9 Truncated TINTIN Construct Purification

Figures 3.9 and 3.10 display the result of test expressions/purifications of truncated TINTIN constructs, that had been designed to identify a stable core of the complex and also to identify boundaries for complex stability. Figure 3.9. displays purifications of constructs containing Eaf7 truncations. These have been grouped into C-terminal truncations and N/C-terminal truncations (88A to 88C and 88D to 88F, respectively).

The majority of C-terminal truncations do not appear to compromise the interaction between Eaf7 and the other TINTIN subunits. In fact with over 50% of the subunit truncated from the C-terminal end (constructs 88A, 88B and 88C), the complex appears to express and purify with similar quality to the full-length complex. EA63B (truncated from residue 143), is the exception to this, as Eaf5 is not obviously present, though Eaf3 and truncated Eaf7 are distinct.

As with full-length TINTIN, all test purifications show similar levels of contamination, likely a consequence of the properties of the resin used. In the case of 88C, intense and smeary bands that do not correlate to TINTIN subunits are also observed, as we might observe in full-length TINTIN purifications.

In contrast to C-terminal truncations, N-terminal truncations up to residue 86 prove problematic for complex integrity. Construct 88D, which also includes a C-terminal truncation from residue 260, is not obviously purified with the Strep-TactinXT resin. Construct 88E (containing residues 86-354) indeed appears to produce Eaf7 co-purified with Eaf3, but Eaf5 is not observed in this purification.

In the case of construct 88F (86-425), the tagged protein is purified and the degradation product of Eaf3 may be present, though the band intensity is weak, and there is no indication of Eaf5.

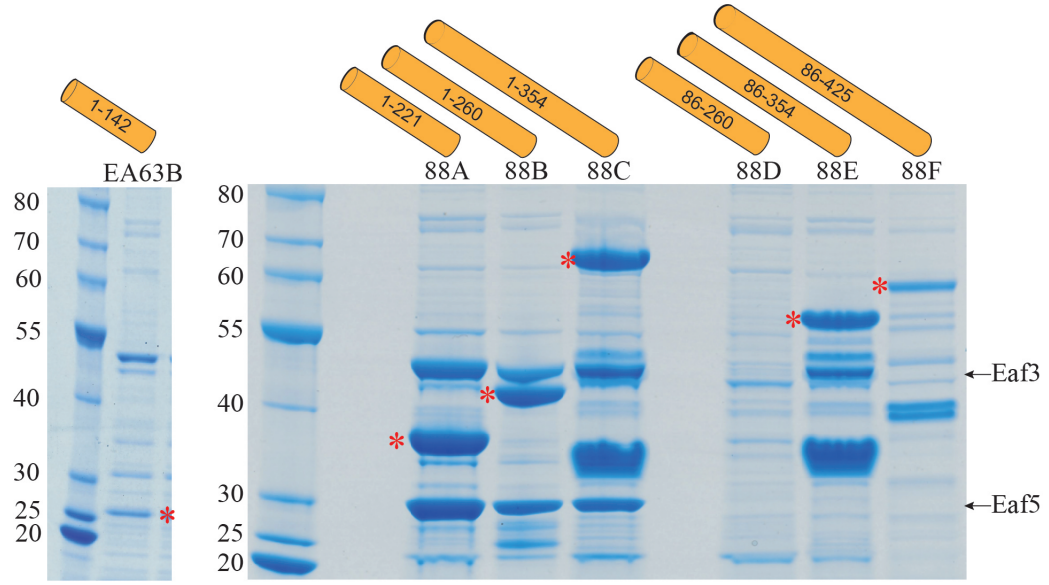

Figure 3.9: Test Strep purification of TINTIN constructs with truncated Eaf7. Proposed Eaf7 bands are marked with an asterisk, and Eaf7 boundaries are represented schematically above the gel. Lanes are grouped between C-terminal and N/C-terminal truncations. C-terminal truncations do not appear to impede complex integrity, as Eaf5 and Eaf3 co-purify with the tagged protein in each case. 88D fails to clearly yield any of the TINTIN subunits, whereas 88E yields bands corresponding to truncated Eaf7, Eaf3 and another intense band that does not correspond to Eaf5. 88F yields an ambiguous result, with a band that potentially corresponds to Eaf7 and potentially a degradation product of Eaf3. Eaf5 does not appear to be present in N-terminal truncated Eaf7 purifications.

Figure 3.10 displays the test purifications of Eaf3/Eaf7 dimers, as well as TINTIN constructs with Eaf5 N-terminal truncations (Eaf7 is C-terminally truncated in each of these constructs; 1-221 and 1-260). These truncations were found to be sufficient for complex integrity in the presence of full-length Eaf3 and Eaf5.

Removal of Eaf5 in combination with Eaf7 C-terminal truncations, appears to severely disrupt complex integrity, as co-purification of full-length Eaf3 via the tagged Eaf7 was not possible (**Figure 3.10A**).

Eaf5 N-terminal truncations also seem to disrupt the yield and integrity of purified TINTIN to a lesser extent (**Figure 3.10B.**). When combined with the smaller of the Eaf7 C-terminal truncations (1-221), it is not clear that a stable complex is present, even at low concentrations. For the construct containing the more complete Eaf7(1-260), Eaf3 co-purifies, though the Eaf5 subunit appears to be less abundant.

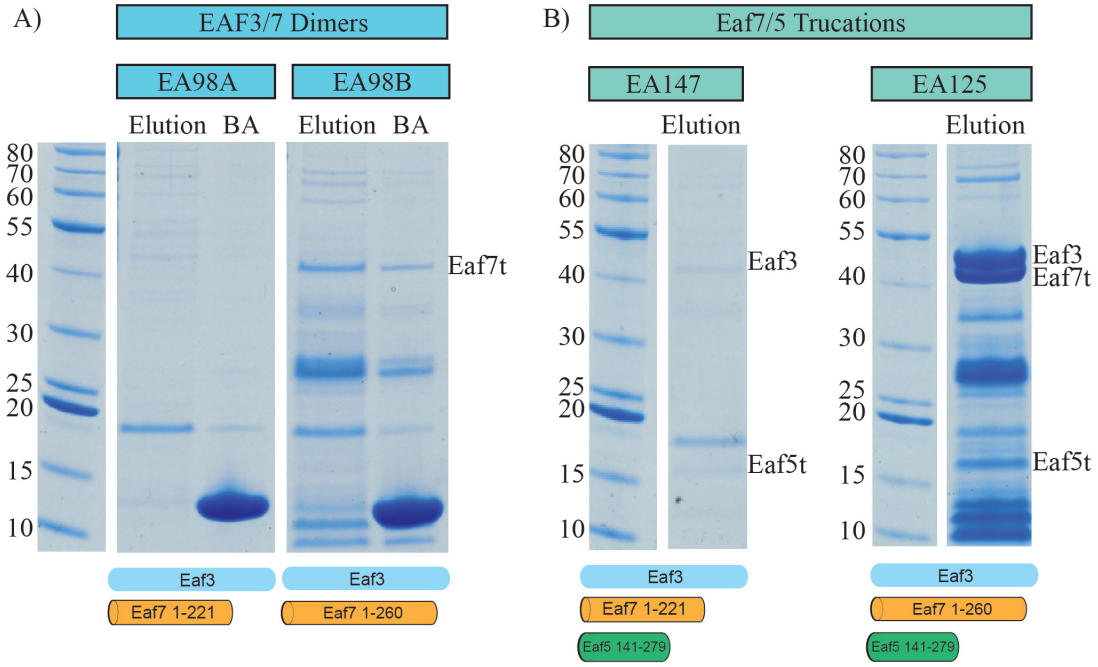

Figure 3.10: Test purifications of Eaf3/Eaf7 dimers (with Eaf7 truncated in each case) and TINTIN constructs containing truncated Eaf7 and Eaf5. Elution = 20 μL sample (from 20 μL + 4 μL loading dye), BA = 10 μL XT resin beads after elution. Schematic representations of each subunit (and residue boundaries) are listed beneath each gel. **A)** Test purification of Eaf3/Eaf7 dimers. 98A purification appears to fail to produce either subunit, while the larger 98B construct yields the truncated Eaf7 band (identical to the band of construct 88B in figure 3.9) along with several other bands that may be degradation products of Eaf7 or Eaf3. **B)** Test purification of the construct EA147 (Eaf7 1-221, Eaf5 141-279), the result of which remains ambiguous, due to several low-intensity bands. In the test purification of construct EA125 (Eaf7 1-260, Eaf5 141-279), the tagged Eaf7 truncation and Eaf3 are clearly present. A band that corresponds with truncated Eaf5 is also observed, this however is less intense than the other TINTIN subunits.

3.10 Upscaling Purification Truncated Constructs

Test purifying the multiple truncated constructs and Eaf3/Eaf7 dimers identified two particularly promising candidates for purification and crystallisation trials. Constructs EA88A and 88B (containing Eaf7 1-221 and Eaf7 1-260, respectively) were expressed in 12L LB (using the LB2 expression scheme described in methods) and subjected to purifications that followed the same general protocol as full-length TINTIN. The result of both sets of purifications can be seen in **Figures 3.11 and 3.12**, for constructs 88A and 88B respectively.

Both constructs appear to express and purify well, with 3 bands corresponding to the molecular weight of each subunit of the TINTIN complex. Early test purifications of the complex suggested that the sample was saturating the Strep-TactinXT resin (using the same volume of resin; 3 mL). As a result, a high-capacity variant of the resin was used, with over 40 mg of protein (determined by bradford assay) detected in elutions.

In the case of EA88A, the protein content was so abundant, that a single IEX chromatography step with the Mono S 5/50 GL could not capture all of the protein. As a result, the run was repeated to capture and further purify the complex.

As with the full-length TINTIN complex, the final size-exclusion step of the purification of either construct, revealed that the protein eluted early for a species of its predicted molecular weight (**Appendix C**).

Generally, both constructs express better in LB than full-length TINTIN expressed in auto-induction media, in terms of mass of pellet to total protein yield. However, both constructs share a similar bottleneck to the final yield of the purified protein, that being the concentration step of the post-SEC sample. As with the full-length construct, when TINTIN is concentrated to > 4 mg/mL, then we observe rapid loss of sample in the region ~50%, depending on the final concentration reached.

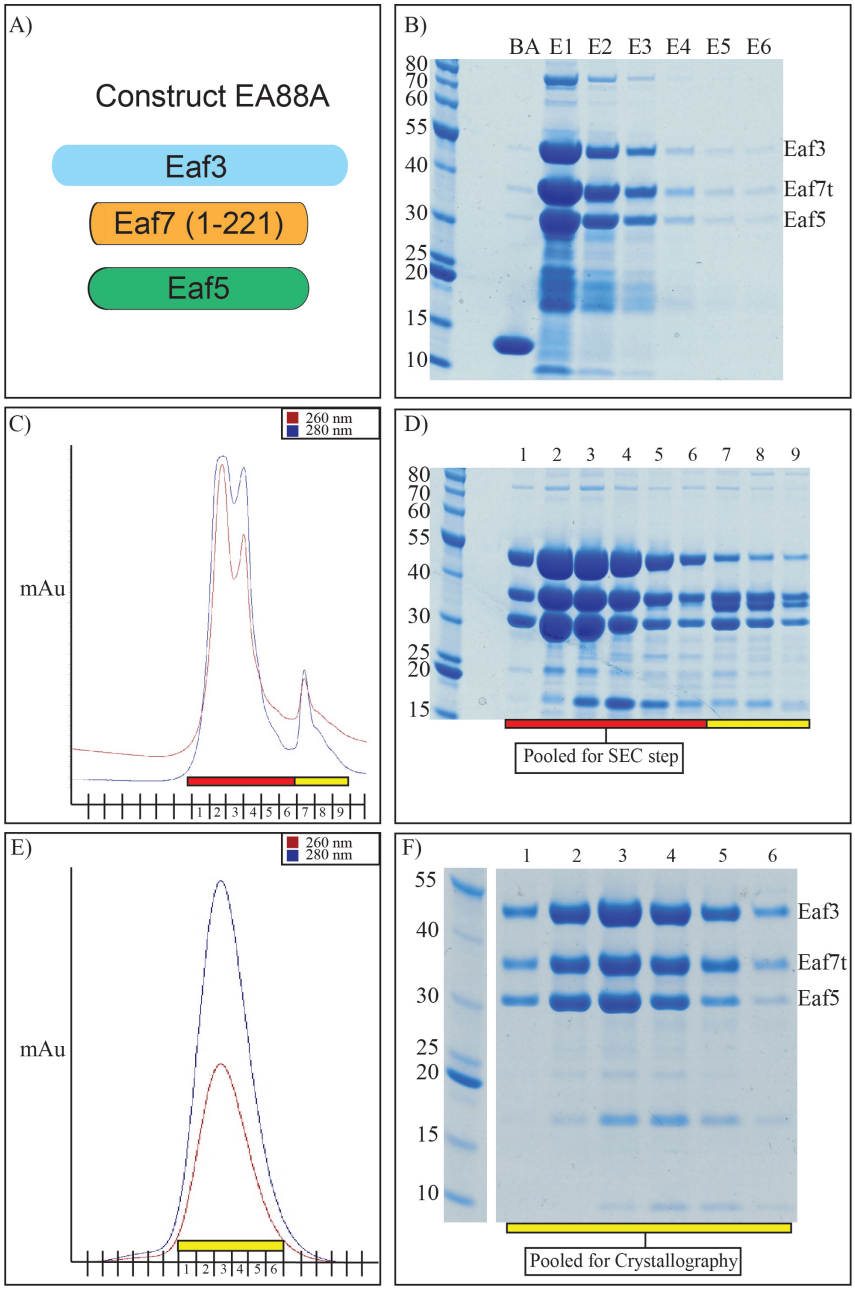

Figure 3.11: Summary of the upscale purification of construct EA88A. **A)** Construct schematic. **B)** Strep-TactinXT high-capacity purification stage. BA = resin after elution of TINTIN with 50 mM biotin (E1–>6). **C)** Ion-exchange step chromatogram (Mono S 5/50 GL). Red and yellow bands correspond to lanes in the adjacent SDS-PAGE gel. **D)** SDS-PAGE analysis of IEX stage of purification, lanes have been colour coded with a red and yellow band to correspond to adjacent chromatogram. Fractions within the red band were pooled for further concentration. **E)** Chromatogram of final SEC step of the purification, using a HiLoad 16/600 Superdex 200 pg. The yellow band beneath the trace corresponds to the yellow band in the adjacent gel. **F)** SDS-PAGE analysis of the final size-exclusion step. Lanes within the yellow band correspond to the yellow band in the adjacent chromatogram and were pooled and concentrated for crystallisation trials.

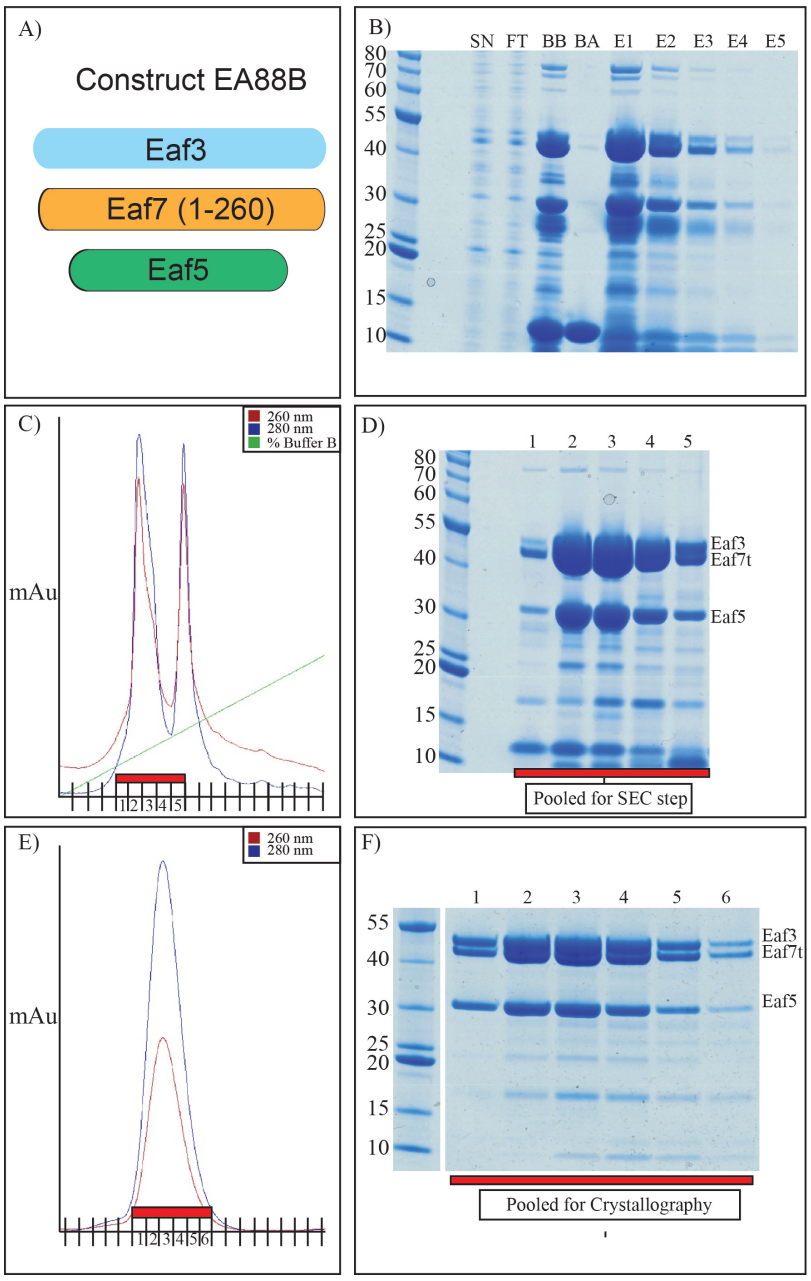

Figure 3.12: Summary of the upscale purification of construct EA88B. **A)** Construct schematic. **B)** Strep-TactinXT high-capacity purification stage. SN = clarified lysate. FT = flow-through after loading the sample on the the Strep resin. BB/ BA = resin before and after elution of TINTIN with 50 mM biotin, respectively. E1-E5 = 2 mL elutions after resuspension step. **C)** Ion-exchange step with a Mono S 5/50 GL. The red band corresponds to lanes in the adjacent gel. **D)** SDS-PAGE analysis of IEX stage of purification. Fractions that correspond to these lanes within the red band were pooled for further concentration. **E)** Chromatogram of final SEC step of the purification using a HiLoad 16/600 Superdex 200 pg. The red band beneath the trace corresponds to the red band in the adjacent gel. **F)** SDS-PAGE analysis of the final size-exclusion step. Lanes within the red band were pooled and concentrated for crystallisation trials.

It was noted that a 15 kDa contaminant band was present in the final purification of the full-length TINTIN complex, as well as both of these truncated constructs. This band was excised and sent for identification to the St. Andrews mass spectrometry service. This determined that the band was a degradation product of Eaf5, more specifically its N-terminal (residues 1-145), that seems to follow the complex throughout the purification, despite multiple rounds of chromatography.

3.11 TINTIN Interaction Studies

The exact biological role of TINTIN is poorly understood and a better understanding of what proteins the complex interacts with, may provide insights into this role. Pull-down assays were conducted between purified TINTIN constructs and peptides that mimic histone N-terminal tails with different methylation patterns, and the RNA Pol II C-terminal domain with different phosphorylation patterns.

Three constructs of increasing complexity were used throughout these pull-down assays. The C-terminal chromodomain domain of Eaf3, and two TINTIN constructs containing a truncated Eaf7 subunit of increasing length; EA88A (Eaf7, 1-221) and EA88B (Eaf7, 1-260) (**Figure 3.13 and 3.14**).

Peptides that had been acquired for these experiments contained biotin tags in order to fix to a Strep-Tactin resin. All proteins used in this project were purified via C-terminal Strep-tags, these were removed via 3C protease that acted on a 3C cleavage site upstream the affinity tag. The process of cleaving the Strep-Tactin, followed by removal of the cleaved tag and protease was a costly process, resulting in loss of over 50% of the sample.

Generally speaking, the affinity between Eaf3 chromodomain and the histone peptides appears to be very weak (**Figure 3.13**).

Comparing the intensity of the cleaved lane (1 mg/mL if protein) with the lanes corresponding to the protein incubated with peptides, shows that most of the protein applied to the resin has not bound. Faint chromodomain bands can be seen in the histone peptide lanes, whereas the CTD peptide lanes appear completely clear. There does not appear to be any visible difference in intensity between bands observed in the histone peptide lanes.

Figure 3.14A and 3.14B shows the result of pull-down assays with constructs EA88A and EA88B (respectively) and the spectrum of peptides tested. Figure 3.14B includes an additional peptide, mimicking the N-terminal tail of H4 (non-methylated). Similar to the chromodomain, the two TINTIN constructs containing C-terminally truncated Eaf7 of varying lengths show weak affinity for histone peptides, irrespective of methylation state (**Figure 3.14**).

The shorter EA88A (Eaf7, 1-221) shows no indication of interaction with CTD peptides, regardless of phosphorylation patterns. The longer EA88B (Eaf7, 1-260) may appear in low-abundance in Pol II CTD lanes, though this is not conclusive.

Comparing the intensity of TINTIN bands between the cleaved lane (that represents the protein at its loaded concentration during the assay) and the assay lanes, suggests that there is weak affinity between TINTIN and the histone peptides. The lack of TINTIN bands in the negative control lane, suggests that there is minimal non-specific binding between the protein and the resin.

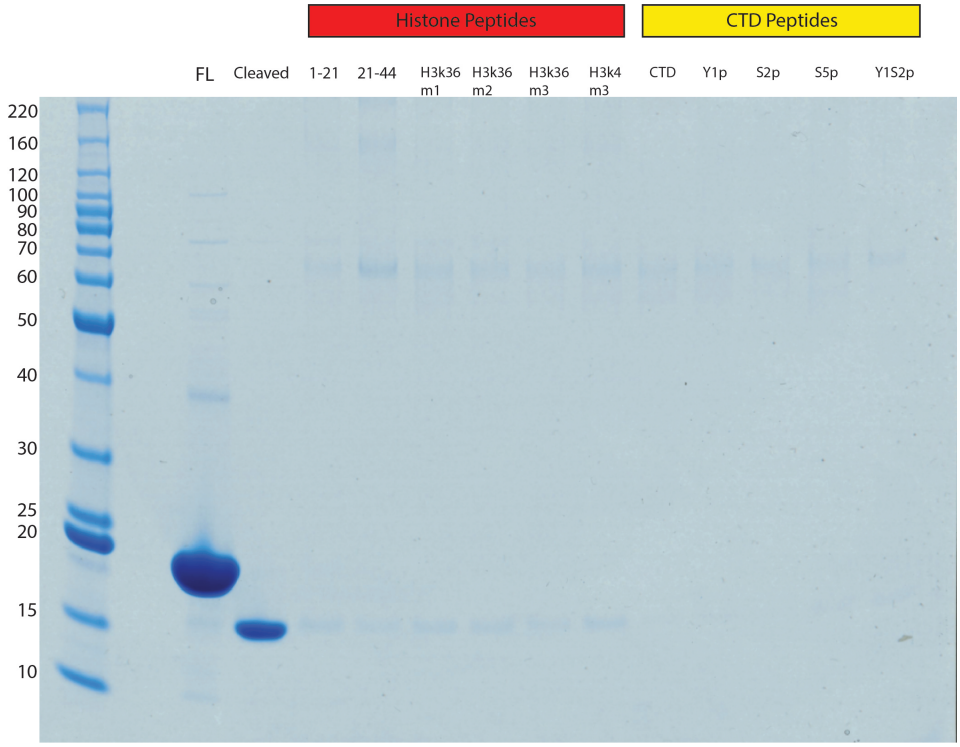

Figure 3.13: Pull-down assay between purified Eaf3 chromodomain and commercially produced peptides, mimicking histone N-terminal tails and Pol II C-terminal domain with varying post-translational modifications. Full-length (FL) and cleaved lanes compare the purified protein to the protein post cleavage step described in methods. The cleaved protein was diluted to 1 mg/mL, 2 μL of this mixture was taken and mixed with 0.4 μL of 6X loading dye, with 2 μL of this loaded onto the gel. Histone peptide lanes labelled 1-21 and 21-44 represent different regions of the unmodified N-terminal tails of H3. Bands in line with the 15 kDa mark represent the cleaved chromodomain. The assay involved the loading of 1 mg/mL of protein onto 20 μL of dynabead Strep-resin. Comparing the intensity of chromodomain bands in the assay lanes to the cleaved lane, show that while some binding to the peptide is observed, the majority of the loaded sample has not bound to the resin. Lanes corresponding to the Pol II CTD with varying states of phosphorylation, do not appear to contain the chromodomain band.

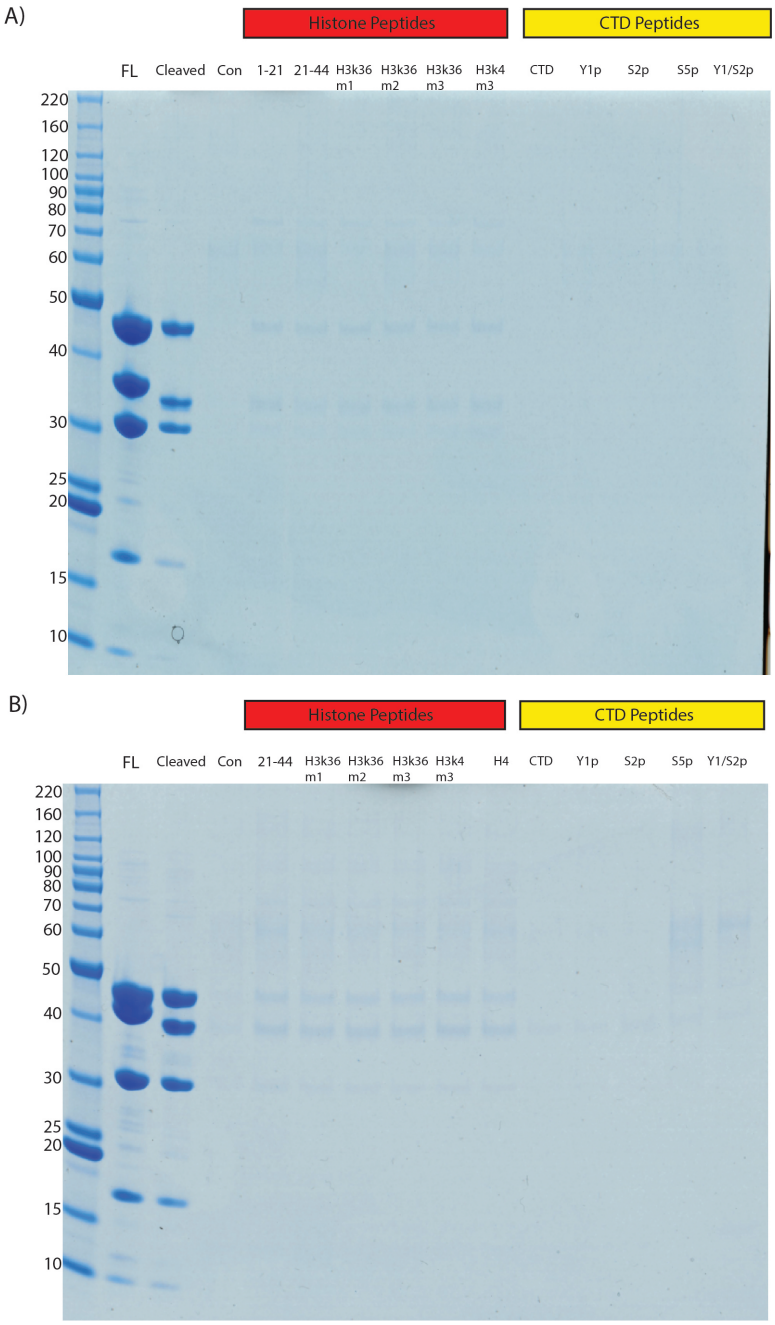

Figure 3.14: SDS-PAGE analysis of pull-down assays between two TINTIN constructs with histone and CTD peptides. "Con" lanes are negative controls, where tag-cleaved protein was mixed with Strep resin. **A)** Pull-down assays using the EA88A construct (Eaf7, 1-221). Faint TINTIN subunits are observed in the histone peptide lanes, though these are low-intensity bands. The Pol II CTD lanes appear completely clear, suggesting no/low-levels of binding occurred between the complex and these peptides. **B)** Pull-down assays using the EA88B construct (Eaf7, 1-260). TINTIN bands appear in the lanes corresponding to the histone peptides (again at low-levels). There is some indication of TINTIN peptides present in the Pol II CTD lanes, suggesting there may be some weak interaction between the complex and Pol II CTD, however these bands are particularly faint and may be a consequence of some non-specific binding.

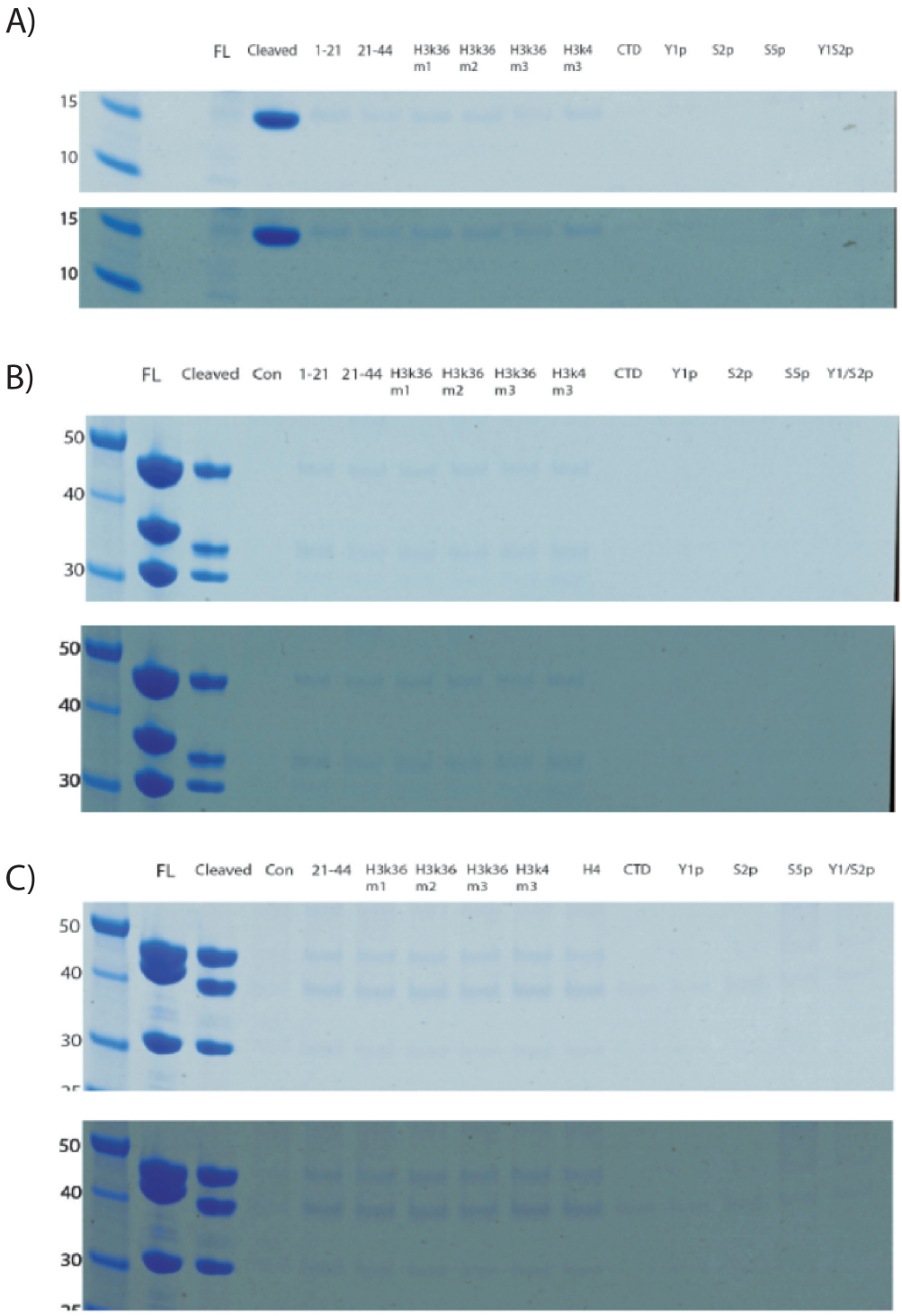

Figure 3.15: Pull-down assays summary, displaying scanned gels above a contrast adjusted image for better visualisation of TINTIN bands. **A)** Eaf3 chromodomain results. **B)** Construct EA88A pull-down results. **C)** Construct EA88B pull-down results.

Chapter 4

TINTIN Purification and Crystallography: Discussion

4.1 TINTIN Expression

While auto-induction media based expressions of TINTIN produce high levels of the complex, this expression proved inconsistent despite multiple attempts to make a more robust expression protocol.

Rosetta 2 cells have been shown to express several of the truncated TINTIN constructs without issue, and cells grow at a consistent rate, irrespective of vectors they are transformed with, though some subtle issue with the cells may be present.

TINTIN appears to express consistently in other media (such as TB and LB), though at significantly lower concentrations. This suggests that both the cells and the TINTIN pET-28a vector is functioning as intended, and the issue is a consequence of the AI media driven expression itself.

Removal of the C-terminal half of Eaf7 and expressing these constructs in LB media circumvented this characteristic inconsistency, which may have been a result of toxicity. Considering the disordered nature of the Eaf7 C-terminal (**Appendix A**), the truncated constructs were favoured for further investigation in the latter half of this project.

4.2 Crystallography

4.2.1 Full-Length TINTIN

Despite inconsistent levels of expression, the full-length TINTIN complex has been concentrated to relatively high levels and subjected to ~3000 crystallisation conditions. Early attempts at crystallisation of the complex produced a majority of completely clear wells in the commercial screens, indicating that protein concentration was well below the threshold for crystal formation.

The optimisation of TINTIN purifications resulted in a 3X increase in the yield of purified complex, to a final concentration of 12.4 mg/mL of protein, though limitations in the final concentration step prevented attainment of higher concentration samples. With increasing concentrations of protein, indication of protein supersaturation was observed. Lowering the temperature that the sample was stored in, from 16 °C to 4 °C also appeared to increase levels of supersaturation.

The majority of wells that showed supersaturation (though still a minority), contained heavily precipitated protein, indicating that either precipitant or protein (or both) concentrations were too high. Preliminary thermofluor assays suggested that TINTIN stabilised in lower salt concentrations and reducing NaCl concentration in the purification buffer saw marginal increases to yield, but did not improve crystallography attempts (**Appendix B**).

While SEC-MALS data (discussed in the next chapter) suggested that the full-length purified sample was close to homogeneity, there were indications that trace amounts of a higher order aggregate were present in the mixture and SDS-PAGE analysis suggested the presence of low MW contaminants (such as a band that consistently runs close to 15 kDa on the gel, that corresponds the C-terminal of Eaf5).

Moreover, disorder prediction software indicates that over 50% of the Eaf7 subunit was disordered from the C-terminal end. It is likely that the combination of low-concentration contaminants, as well as some intrinsic disorder within the complex, prevent the nucleation of protein crystals and further biophysical analysis and truncations (of TINTIN subunits) were carried out to circumvent this issue.

These truncations would also serve to identify regions that were essential for complex assembly and function (in later pull-down assays).

4.2.2 Truncated Constructs

Removal of C-terminal disordered region from Eaf7, dramatically improves protein yield per gram of cell pellet used. Both EA88A (Eaf7, 1-221) and the larger EA88B (Eaf7, 1-260) express at consistently high-yields, when compared to the full-length complex.

An encouraging indication of improvement in complex stability, was the absence of high-intensity Eaf7 degradation products in the early stages of the purification, that were typical of purification of the full-length complex.

Despite further crystallisation trials, neither of these complexes produced crystals, though showed a higher proportion of supersaturation in the wells of the commercial screens. This effect was more pronounced when screens were stored at 4 °C.

Higher concentrations of protein were not tested (>8 mg/mL) for two reasons. Firstly, the majority of supersaturated wells contained precipitated protein and secondly, both complexes suffered from the same loss of sample during concentration, as the full-length complex. The latter, limited the number of screens that could be tested with every purification attempt.

It has been proposed that the interaction between Eaf5 and the rest of the TINTIN complex is relatively weak. *In vitro*, this weak interaction could translate to an equilibrium between the complete trimer and an Eaf5 subunit free of an Eaf3/Eaf7 dimer. This equilibrium would result in polydispresity of the species in solution, which is not conducive to crystal formation.

Interestingly, low-MW contaminants are present in the final sample and may be disrupting crystal formation, by increasing polydispersity. This includes the Eaf5 C-terminal, which is essential for interaction with Eaf7 and the rest of NuA4 (Setiaputra et al. 2018).

4.3 TINTIN Architecture

The pursuit of a more stable TINTIN trimer, has provided insights into the overall architecture of the complex (via disorder prediction, limited proteolysis and subsequent generation of truncated constructs).

4.3.1 The Role of Eaf7

Firstly, we can clearly see that predicted disordered C-terminal region of Eaf7 is dispensable for complex integrity, provided both Eaf5 and Eaf3 are present. 3 C-terminal truncations of Eaf7: EA88A(1-221), EA88B(1-260), EA88C (1-354) appear to co-purify Eaf5 and Eaf3. However, the greatest C-terminal truncation (Eaf7 1-142) shows weaker affinity for Eaf5, suggesting some significant interaction occurs between Eaf5 and residues 142-221 in Eaf7.

When C-terminal truncations are combined with truncations in the N-terminal region (which disorder prediction software suggests is stable), the intensity of TINTIN bands reaches background levels, or seems to alter the stoichiometry of subunits. EA88D (Eaf7, 86-260) does not obviously produce any of the TINTIN subunits, highlighting the importance of the N-terminal for complex assembly. However, the EA88E complex (Eaf7, 86-354) co-purifies Eaf3, suggesting additional points of contact may exist between Eaf3 and Eaf7's C-terminal.

4.3.2 Eaf3/Eaf7 Dimers

The human orthologs of the Eaf3 and Eaf7 subunits form a stable dimer *in vivo* (Xie et al. 2015). Whether or not this is reflected in yeast has yet to be verified. Two constructs of the dimer were generated (EA98A and EA98B) and test expressed/purified, both containing the MRG and MBD domains found to the basis for the assembly of MRG15/MRGBP complex.

Both constructs contained truncations of the Eaf7 subunit (Eaf7, 1-221 and 1-260). These truncations had not disrupted complex formation, when Eaf5 was also expressed (EA88A and EA88B respectively).

Interestingly, the absence of Eaf5, in combination with these C-terminal truncations appears to completely disrupt complex formation. In the case of the smaller Eaf7 construct (1-221), the tagged protein is not purified. From this, it may be inferred that Eaf5 plays a role in complex integrity and Eaf7 stability is also dependent on interaction with the other subunits.

A full-length Eaf7 may restore stability and result in co-purification of the Eaf3 subunit, considering that Eaf7 constructs containing residues 86-354 appears to purify Eaf3. Nevertheless, current *in vitro* data, indicates that Eaf5 contributes to Eaf3/Eaf7 interactions.

4.3.3 Eaf5 Truncations

Purification of full-length TINTIN and constructs with Eaf7 truncations, consistently produce a low-MW contaminant band that was found by mass spectrometry analysis to correspond to the C-terminal of Eaf5. Interestingly, Setiaputra et al. (2018) cross-linking mass spectrometry data, suggests that this region makes contact with Eaf7 and the Eaf1 scaffold of NuA4.

Two constructs containing both Eaf7 truncations and an Eaf5 N-terminal truncation (Eaf5, 141-279) were generated, test expressed and purified. As with the Eaf3/Eaf7 dimer investigation, the Eaf7 truncations included, were not observed to disrupt complex formation, when both full-length Eaf3 and Eaf5 subunits were present.

N-terminal truncation of the Eaf5 subunit can be sufficient for complex stability, provided the larger Eaf7 truncated construct is present (1-260). The presence of Eaf5 C-terminal clearly restores complex stability, possibly due to interactions with Eaf7 and Eaf3 in this region.

The EA125 construct, containing the full-length Eaf3, Eaf7 (1-260) and Eaf5 (141-279), represents the smallest construct of the TINTIN complex thus far. Intensity of bands in SDS-PAGE gels, indicate that the Eaf5 subunit is less abundant than the rest of the complex, possibly due to a reduction in strength of interaction. A map of possible interaction interfaces between TINTIN subunits is included in chapter 7, including residue boundaries for positions of interaction.

4.4 TINTIN Interaction Studies

TINTIN's interaction profile is poorly characterised and would likely help explain its biological role. Eaf3 contains a chromodomain responsible for interacting with trimethylated H3K36, though the strength of this interaction was found to be relatively weak (Sun et al. 2008). However, these experiments did not include the complete Eaf3 protein, or Eaf3 in complex with the remaining TINTIN subunits.

Moreover Rossetto et al. (2014) have proposed that the complex directly interacts with the elongating Pol II via its C-terminal domain, though direct evidence of this has yet to be produced. To that end, the long-form Eaf3 chromodomain (Sun et al. 2008) and truncated TINTIN constructs EA88A and EA88B, were subjected to pull-down assays with peptides that mimic histone N-terminal tails and Pol II CTD at different stages of transcription.

4.4.1 Histone Peptide Interactions

For all 3 proteins tested, the majority of purified sample does not bind to any of the peptides. This is evident when we compare the intensity of "cleaved" lanes, which are representative of the quantity of sample exposed to resin-fixed peptide. Faint bands suggest that any interaction observed, is weak.

The presence of the complete Eaf3 (and additional TINTIN subunits) does not obviously improve affinity between the chromodomain and H3K36me3. Though a more sensitive and precise experimental technique such as BLI may reveal differences in affinity. Moreover, no obvious difference in band intensity between histone peptides is observed, as a result we cannot exclude the possibility that these faint bands are a consequence of non-specific binding, though controls suggest otherwise.

4.4.2 Pol II CTD Interactions

The chromodomain was not expected to bind to Pol II CTD and pull-down assays show that while the domain non-specifically binds the histone peptides, CTD lanes appear completely clear.

For the smaller of the two TINTIN constructs; EA88A (Eaf7 1-221), CTD lanes appear to be completely clear. Conversely, for the larger TINTIN complex (Eaf7 1-260), faint bands of TINTIN are present in the CTD lanes. While noticeably less intense than bands in the histone peptide lanes, TINTIN subunits from this construct appear in all of the CTD lanes.

The additional 22 residues (between EA88B and EA88A) may confer affinity for Pol II CTD, though it must be noted that this seems to be non-specific, as even the non-phosphorylated CTD appears to be bound. Whether this interaction is due to some specific interaction is not obvious from this data. Furthermore, the results are unable to determine differences in affinity for various phosphorylation patterns, a limitation of evaluating these interactions by SDS-PAGE.

However, the fact that a more complete TINTIN complex shows signs of interaction with Pol II CTD is interesting. A more sensitive technique with increasingly complete TINTIN complexes, may validate the theory of direct interaction between TINTIN and Pol II CTD.

Chapter 5

Architecture of the TINTIN Complex: Results

5.1 Summary

The crystallisation trials of the full-length TINTIN complex (and truncated constructs) failed to produce crystals. The monodispersity of the sample (and the truncated EA88B) was assessed by SEC-MALS, which also gave an indication of the oligomeric state of the sample. This data shows that TINTIN existed as a single trimer, though trace amounts of higher order aggregates may be present.

A truncated TINTIN complex (EA88B) was subjected to native ion mobility mass spectrometry, to identify contaminant proteins and determine the presence of flexibility within the complex. From this data, we verified the presence of a monomeric TINTIN species that, once isolated via quadrupole and subjected to collision induced dissociation, produced a Eaf3/Eaf7 dimer and Eaf5 monomer.

TINTIN related species identified in the spectra were then further analysed by ion mobility, suggesting individual subunits are extended. Eaf5 in particular adopts a highly extended conformation, with a high collision cross section and charge, when compared to ordered proteins of comparable mass. The TINTIN complex however, is likely ordered, based on its collision cross section:mass and charge state distribution:mass ratios.

To better understand the structure of the complex in the absence of high-resolution data, SAXS was carried out on the two most successful truncated TINTIN constructs (EA88A and EA88B). Datasets collected, reveal a concentration dependence on the calculated MW, possibly a consequence of oligomerization events. Higher concentration samples however, produce a calculated MW which approximates well with monomeric TINTIN. Both constructs produce similar low-resolution, highly extended structures of comparable dimensions, as was predicted when considering chromatograms from SEC-purification steps.

5.2 Light Scattering Analysis

SEC-MALS

A common feature of TINTIN purifications was the unusually early elution volume for a complex of its predicted molecular weight, from size-exclusion columns (HiLoad Superdex 200 16/600 GL)(**Chapter 3**). According to our own calibrations using MW protein standards (**Appendix C**), the sample was eluting at a volume that would correspond to approximately 3X the theoretical MW of TINTIN (~130 kDa). This suggested that we were observing a 'trimer of trimers' and a similar effect was observed when purifying the truncated TINTIN constructs (EA88A and EA88B).

We employed size-exclusion chromatography multi-angle light scattering (SEC-MALS) to determine an absolute mass of the complex and to get an indication of sample polydispersity for the full-length construct and EA88B (Eaf7, 1-260) (**Figure 5.1 and 5.2**).

Analysis of the full-length complex, reveals the polydispersity of the major peak in Figure 1 (MW/Mn) = 1.007 (± 0.987%), indicating that the sample is close to homogeneity. The light scattering signal, suggests the presence of higher order aggregates distinct from the primary peak, though these probably account for a small fraction of the sample, when considering the low-intensity signals for corresponding refractive index and UV measurements at these positions on the chromatogram.

Molar mass distributions suggest that the full-length TINTIN is 123.7 kDa, which equates roughly to the predicted MW of monomeric TINTIN (theoretical MW of 130 kDa). However, the molar mass distribution forms a slight downward slope within this peak, which leads into a plateau with a MW of 122.5 kDa. This also aligns with the UV/LS tail in the chromatogram.

Similar to the full-length construct, the EA88B construct (Eaf7, 1-260) is observed in its monomeric state (approximately 100 kDa). The complex appears to show close to ideal levels of purity in the primary peak(MW/Mn = 1.003 ± 0.413%), with no obvious sign of higher order aggregates. Overall, these results show that the constructs had been purified in their monomeric forms, though the tails in each trace suggest the presence of a marginally smaller complex.

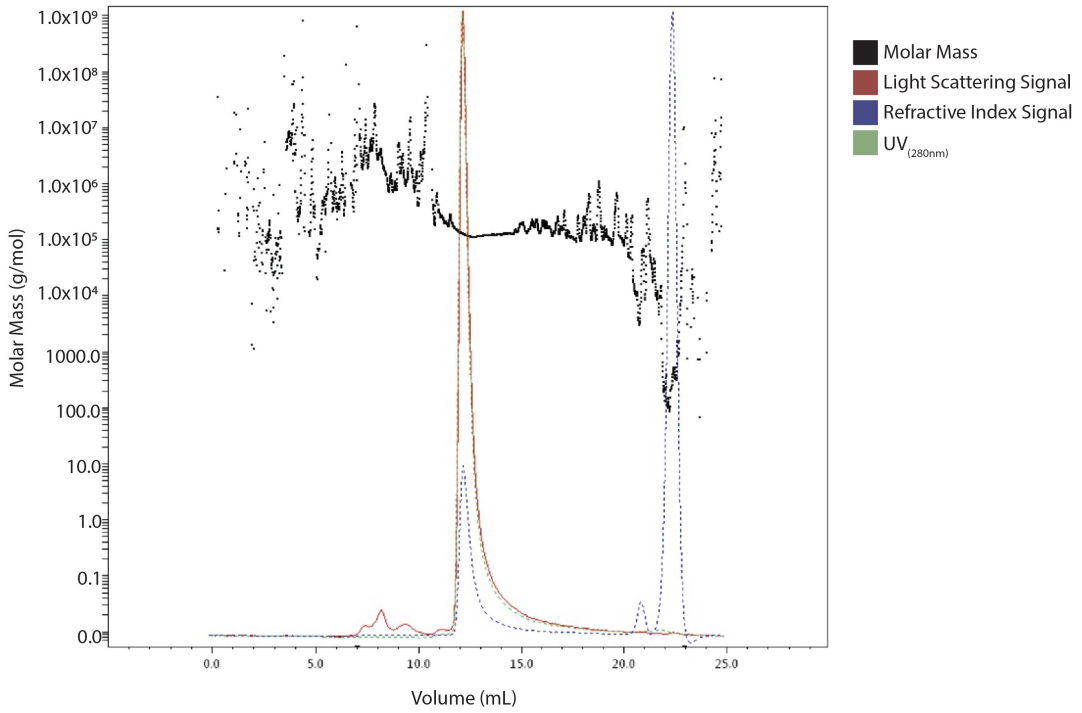

Figure 5.1: Molar Mass distribution from a SEC-MALS experiment. 0.84 mg/mL of purified TINTIN was loaded onto a Superdex 200 increase 10/300 GL, connected to a HPLC system. The sample was loaded and run through the column at a flow-rate of 0.5 mL/min at 25 °C. A Dawn 8+ MALS detector and an Optilab T-rEX refractive-index detector, were used to collect data. Light scattering is displayed in red, refractive-index in blue and UV is displayed in green. A strong UV/LS/RI peak is detected at ~12mL, though there is also a weak LS signal between 7-10 mL. The major peak is followed by a slight tail, but otherwise appears to account for the majority of protein loaded onto the column. Molar mass within the major peak appears to slope downward, before reaching a stable plateau that corresponds to the low-intensity UV/LS/RI tail.

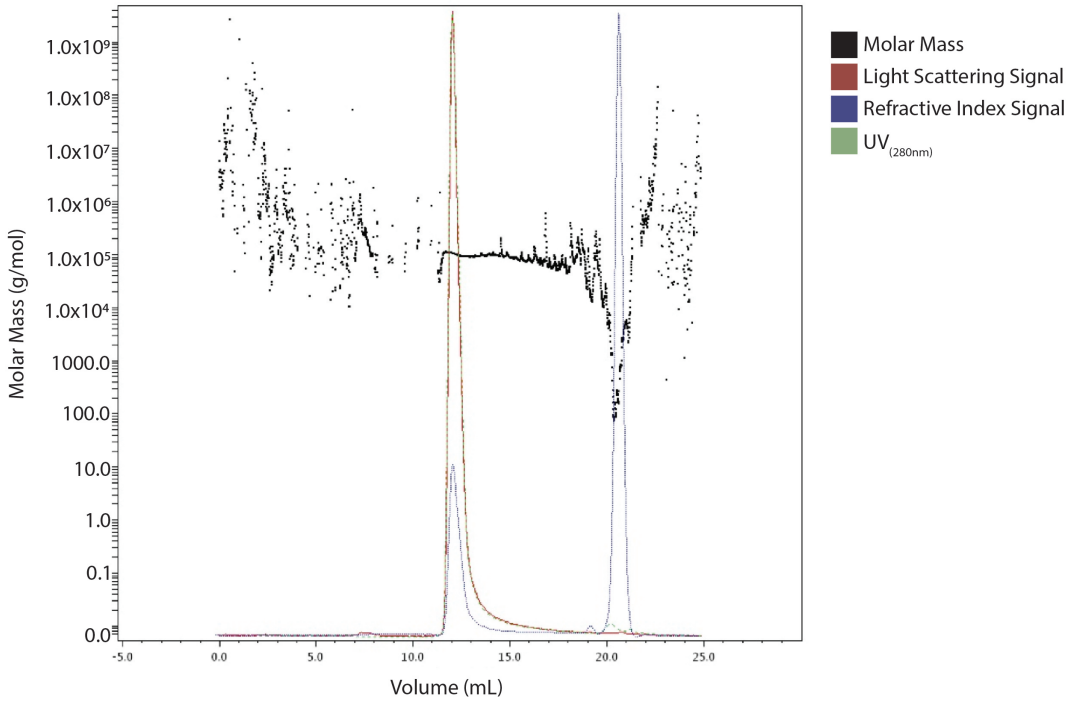

Figure 5.2: Molar Mass distribution from a SEC-MALS experiment of truncated TINTIN construct, EA88B (Eaf7, 1-260). 0.71 mg/mL of purified protein was loaded onto a Superdex 200 increase 10/300 GL, connected to a HPLC system. The sample was loaded and run through the column at a flow-rate of 0.5 mL/min at 25 °C. A Dawn 8+ MALS detector and an Optilab T-rEX refractive-index detector, were used to collect data. Light scattering is displayed in red, refractive-index in blue and UV is displayed in green. As with the full-length complex, a strong UV/LS/RI peak is detected at ~12 mL. This peak is also followed by a slight tail, but otherwise appears to account for the majority of protein loaded onto the column. The molar mass curve within the major peak is not as slanted as the full-length complex.

5.3 Native Mass Spectrometry

5.3.1 Experimental Summary

Native ion mobility mass spectrometry was carried out in the Thalassinos lab with Charles Eldrid, who also carried out analysis of the raw data generated from the following experiments. Data was collected using a Synapt G1 (Waters, UK). Experiments were conducted with the truncated TINTIN construct EA88B, containing a C-terminally truncated Eaf7 (1-260), which was buffer exchanged into 100 mM ammonium acetate solution and diluted to 5-10 μM (0.7 mg/mL TINTIN). The exchange process was costly, with over 50% of the sample being lost as a result. The mass of each species was determined using MassLynx and Amphirite, which was also used to calculate collision cross sections (CCS) of TINTIN and associated species.

5.3.2 Identification of TINTIN Species

The full spectrum of the truncated TINTIN construct can be seen in Figure 5.3. Several species were identified that corresponded to the complete complex, with a calculated MW close to that of the theoretical value (111 kDa). TINTIN was found to adopt 3 charge states (+21 to +23).

Interestingly, other TINTIN related species were also detected besides the complete complex. Most notably, the subunit Eaf5 was detected in 14 different charge states (+11 to + 25), indicative of a disordered protein. The individual Eaf7 subunit was also detected, adopting significantly fewer charge states (+12 to +14), while Eaf3 was not detected as an individual subunit. Futhermore, low-intensity signals were detected for dimeric Eaf5/Eaf3 (+17 and +18) and possibly a dimeric form of the TINTIN complex (+31).

A dimeric complex of Eaf3/Eaf7 was not observed, despite the presence of Eaf5 monomer in the spectra.

Two additional species were identified in the spectrum, with MW not immediately correlated to a combination of TINTIN subunits. An 11 kDa species is found, occupying charge states +6–>+9 and a 91 kDa species, occupying 2 charge states (+20 and +21) are consistently observed in these experiments.

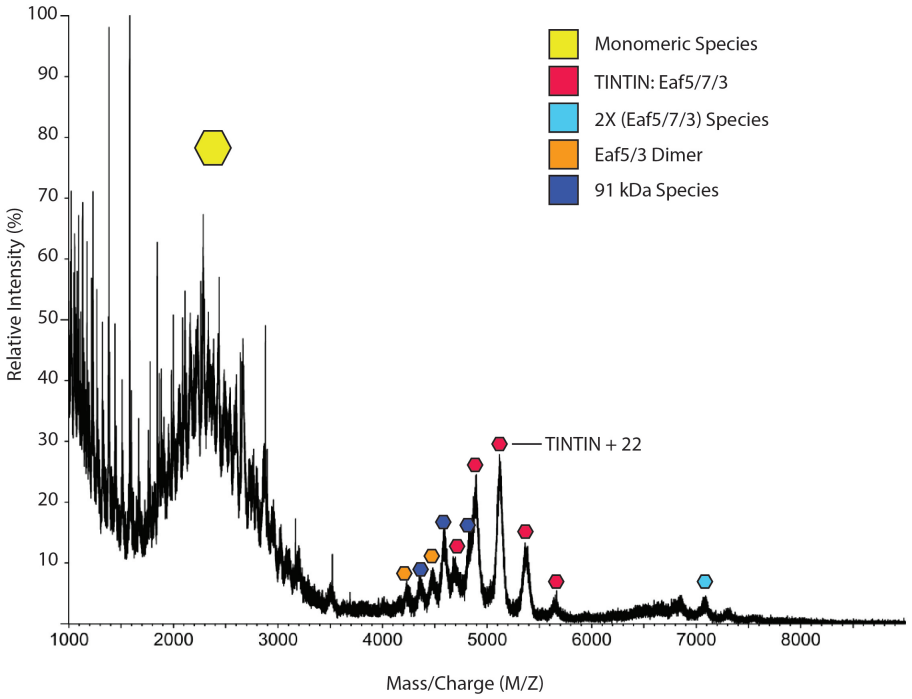

Figure 5.3: Native spectrum of the truncated TINTIN construct (EA88B) from a Synapt G1, plotting the relative intensity of species observed against their mass/charge ratio. Significant peaks (corresponding to TINTIN or its component subunits) have been annotated and colour coded. Species corresponding to the complex were detected. Species that correspond to a dimer of Eaf5/Eaf3 are also observed. A low-intensity signal that may correspond to a dimerization of TINTIN is also detected. Multiple peaks that correspond to Eaf5 and Eaf7 subunits were also detected (the group of peaks have been labelled yellow) with Eaf5 adopting 14 different charge states. A dimer of Eaf3/Eaf7 was predicted to be observed due to the presence of Eaf5 monomer, but was not observed after multiple iterations of the experiment. A species that correspond to the Eaf5/Eaf3 dimer and corresponding Eaf7 are observed in the spectra. A +22 charge state TINTIN species was isolated using the quadrupole for further analysis. Unidentified 11 kDa species (in the monomeric region) and a 91 kDa species are also detected.

5.3.3 Collision Induced Dissociation

The +22 TINTIN complex was isolated by quadrupole and subjected to collision induced dissociation (**Figure 5.4**). The voltage of the collision cell was increased in 10 V steps and upon increasing the trap and transfer to 70 V, a dissociation event appears to occur, generating multiple highly charged species with low-M/Z values, as well as larger species with lower charged states. The smaller species were calculated to have a MW of 31 kDa, whereas the larger species were calculated at 79 kDa, which correspond to Eaf5 and an Eaf3/Eaf7 dimer respectively.

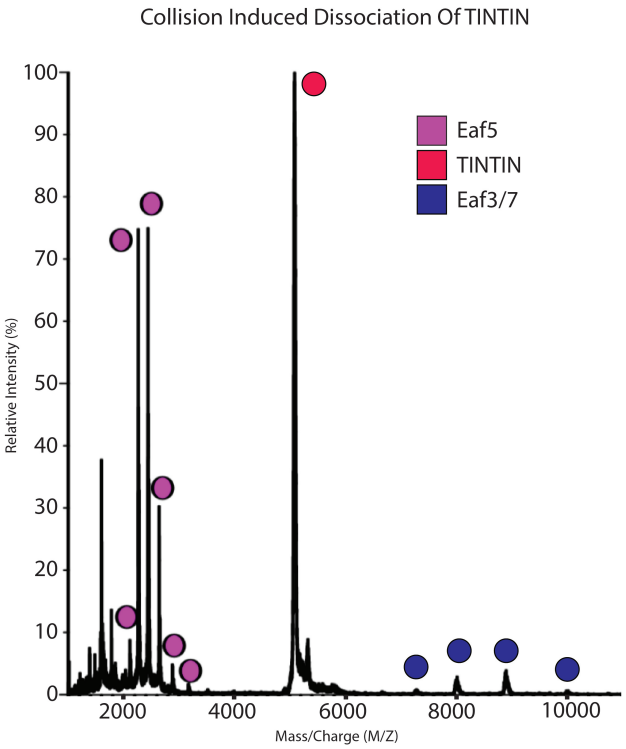

Figure 5.4: Spectrum displaying the collision induced dissociation of the quadrupole isolated +22 charge state TINTIN complex. The spectra is produced when the trap and transfer is increased to 70 V, producing two sets of species in addition to the complete complex. A set of lower M/Z species with a calculated MW of 31 kDa was detected, corresponding to the Eaf5 subunit, these species are shown to contain high charge states for a protein of this MW. Another set of high M/Z species (with low-charge values for species of this size) were detected in parallel with the Eaf5 species. The MW of this larger species was calculated to be approximately 79 kDa, which corresponds to a dimer of Eaf3/Eaf7.

5.3.4 Collision Cross Section Calculation

Native ion mobility was used to determine the collision cross section (CCS) of TINTIN (and associated species), to gain a better understanding of physical dimensions of the complex and to determine regions of flexibility/disorder within the structure. Collision cross section distributions of each of these species, was calculated and assessed, providing several structural insights (**Figure 5.5**).

Eaf5 adopts a broad range of CCS values (2600-5780 $Å^2$) (**Figure 5.5A**). This distribution curve reveals a large CCS peak, beside a smaller CCS peak with a significantly lower intensity. This suggests that the protein is found in two populations, a smaller CCS (more compact) conformation and a larger CCS (more extended) conformation, with a higher population.

Figure 5.5B demonstrates the influence of the complex formation on the folding of individual subunits. The CCS distribution of the Eaf5/Eaf3 dimer, shows a highly populated peak between 4000 and 5000 $Å^2$, followed by a minor population with CCS closer to 6000 $Å^2$. This is similar to what was observed for the individual Eaf5 subunit, though the dimer appears to preferentially adopt a more compact conformation in this instance.

Eaf7 has a comparatively simpler CCS distribution curve (compared to Eaf5 species) (**Figure 5.5C**). The CCS distribution of Eaf7 contains a single, significantly narrower peak. (2840 - 3310 $Å^2$). However, the slight tail in this peak that reaches towards higher CCS values is also noted and is possibly a reflection of some conformational flexibility. Figure 5.5D displays the CCS distribution of the complete complex. The CCS range for TINTIN is a single (relatively narrow) peak, spanning 5680-5740 $Å^2$. The peak in question has a normal distribution, suggesting the complex formation has stabilised the otherwise flexible/disordered subunits.

The two unexpected species (11 kDa and 91 kDa) were also subjected to this analysis (**Figure 5.6**). The 11 kDa species (**Figure 5.6A**) occupies two CCS populations, a less populated, compact form and a highly populated, expanded form, similar in behaviour to Eaf5. The 91 kDa species shows an a preference for a more compact conformation with a small subset in an extended/disordered conformation, as we observe with the Eaf5/Eaf3 dimer (**Figure 5.6B**).

To get a better understanding of the extent of flexibility within the complex and its component subunits, the median CCS and median charge of each species was plot against their MW (**Figure 5.7**). These were compared to mathematical models generated by Beveridge et al. (2014), that are representative of ordered and disordered proteins.

Generally speaking, the majority of TINTIN and associated species detected in the native spectra, appear to possess CCS and charge state distributions, expected of folded proteins of equivalent molecular mass. One notable exception however, is the Eaf5 subunit. Eaf5 adopts many charge states for a relatively small protein. Moreover, as previously mentioned, the CCS values calculated for this monomeric species, are comparable to that of the Eaf5/Eaf3 dimer, as well as the full-length complex.

The unexpected species are also plot against these models, to assess their order/disorder. The 11 kDa species has a CCS:mass ratio, similar to that of a disordered protein. The 91 kDa species however, has a relatively compact conformation for a protein of its MW.

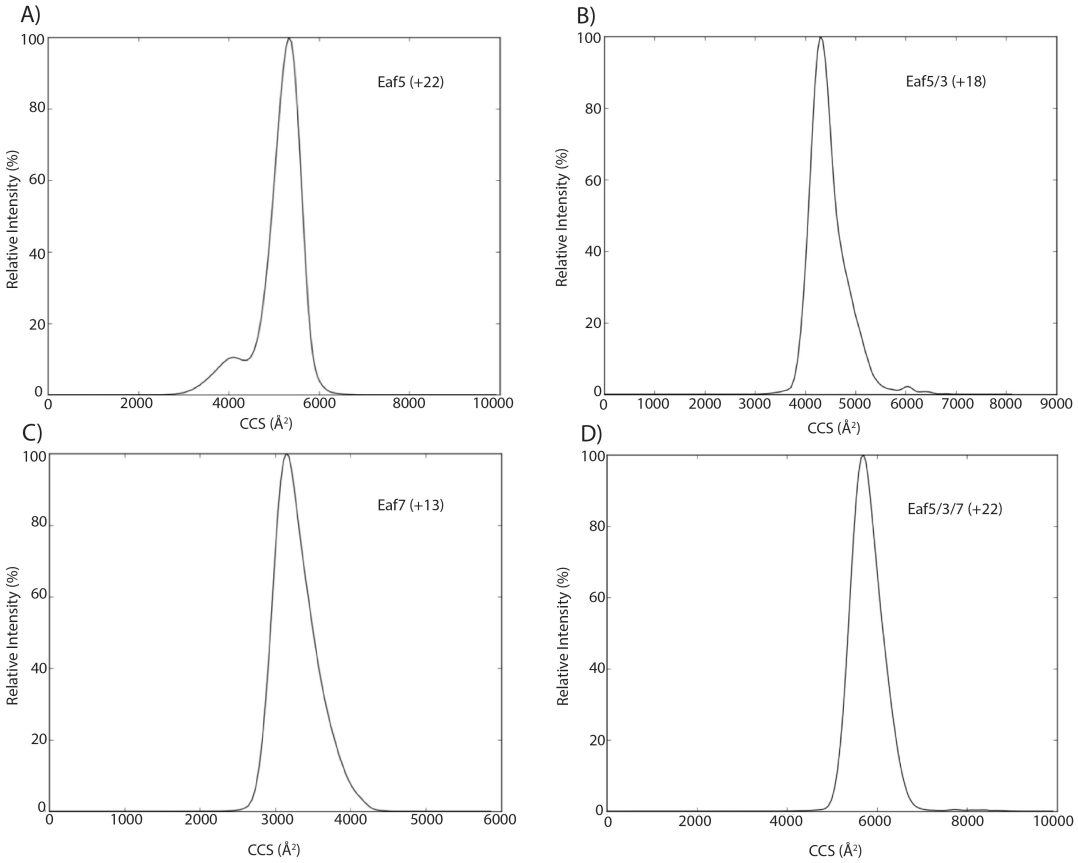

Figure 5.5: Collision cross section distribution curves of the monomeric TINTIN and associated species. Plotting relative intensity (%) against calculated CCS. Charge states are listed in parentheses. **A)** CCS distribution curve of Eaf5. The protein is found in two populations, a more compact (but less abundant) CCS value population ~4000 $Å^2$ and a more extended (and more abundant species) between 5000-6000 $Å^2$. **B)** CCS distribution of the Eaf5/3 dimer. As with the monomeric Eaf5, the dimer appears in two distinct populations, however in this instance, the complex appears to preferentially adopt a more compact/folded conformation. This is possibly indicative of a stabilising affect on the Eaf5 subunit, as a result of associating with Eaf3. **C)** CCS distribution of the Eaf7 subunit. The CCS distribution is narrower compared to Eaf5, though the curve clearly contains a slight tail, possibly a reflection of the subunit's inherent flexibility. **D)** CCS distribution of the complete TINTIN complex. The distribution is uniform and presents a single, narrow population, suggesting that trimer formation stabilises the component subunits.

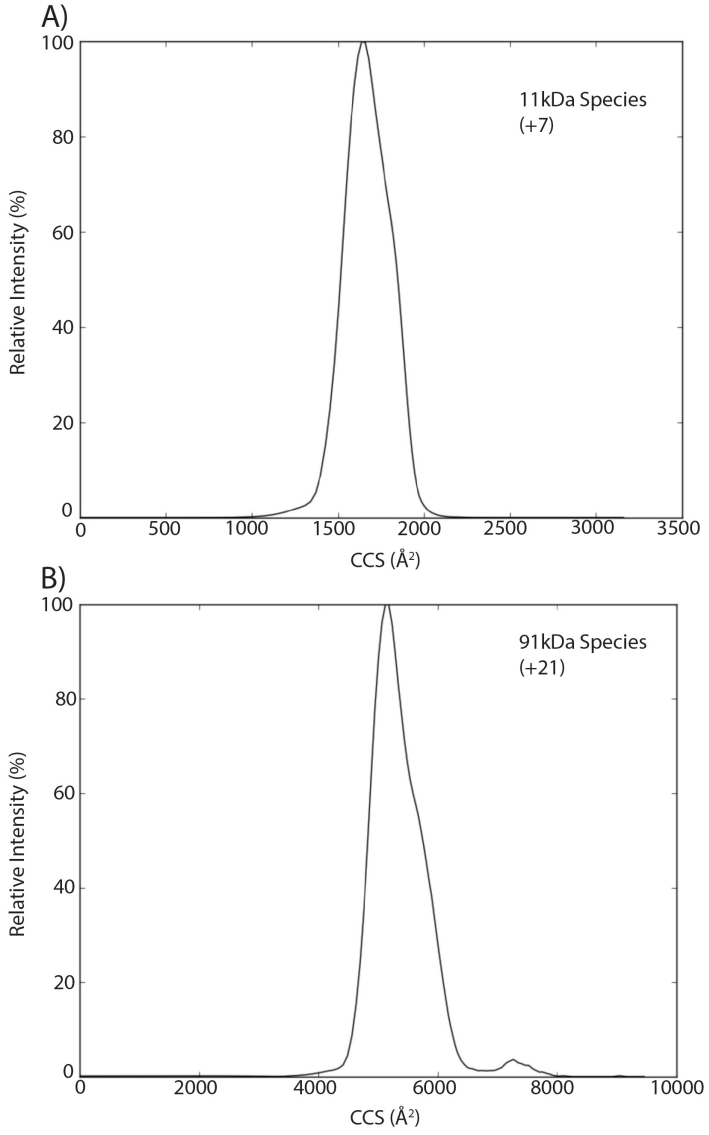

Figure 5.6: Collision cross section distribution curves of species that were not predicted to be present in the TINTIN sample, plotting relative intensity (%) against calculated CCS. **A)** The 11 kDa (+7 species) may adopt two conformations, though the less compact of these conformations is preferentially occupied. **B)** The 91 kDa (+21) species also adopts two conformations, though the more compact conformation is preferentially occupied.

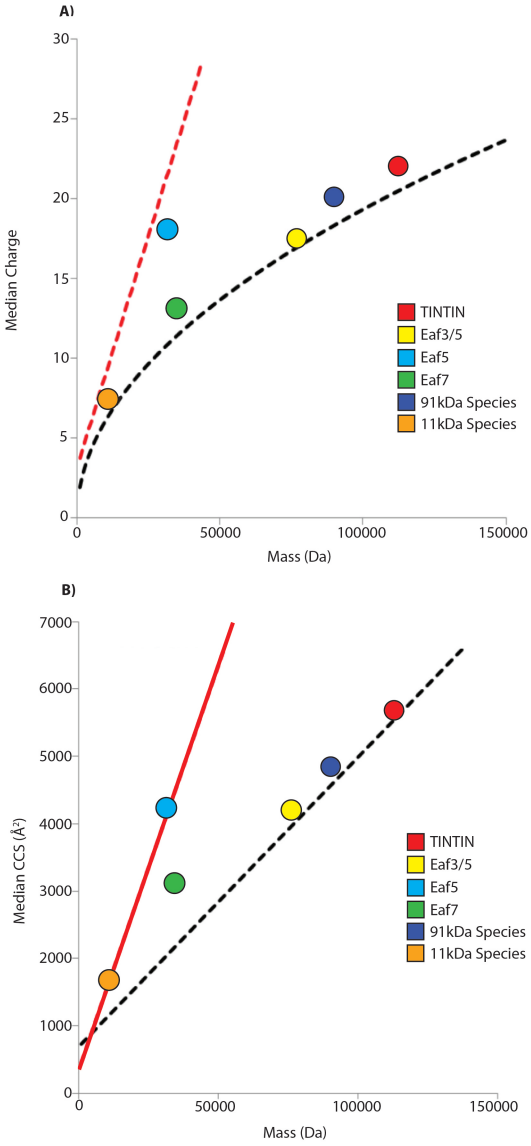

Figure 5.7: Comparison of physical properties between TINTIN (and associated species), against mathematical models generated by Beveridge et al. (2014), representing ordered (black) and disordered (red) proteins. **A)** Median charge against MW of experimental species, these have been annotated and colour coded. TINTIN, the Eaf5/Eaf3 dimer and Eaf7 have charge states indicative of ordered proteins, of equivalent MW, whereas Eaf5 clearly adopts a higher charge state for a relatively small protein, indicative of an extended conformation. **B)** Median CCS vs MW. TINTIN and associated species have been annotated and colour coded. TINTIN and the dimeric Eaf5/Eaf3 show collision cross sections expected of ordered proteins of equivalent mass, while both monomers show high CCS values, for relatively low-MW proteins.

5.3.5 Validation of Eaf5/Eaf3 dimer

The presence of the an Eaf5/Eaf3 dimer in the native ion mobility experiments, was an unexpected result when we consider the data from Rossetto et al. (2014). When considered in parallel with truncation studies mentioned in chapter 3, it was speculated that a direct contact between the two subunits was present within the complex.

As a result, a construct containing full-length Eaf3 and Eaf5 was generated, with a C-terminal Strep-tag on the Eaf3 subunit (**Figure 5.8**). This construct was expressed in Rosetta 2 cells, using LB media at 37 °C until OD reached ~0.6-0.7, before induction with 0.2 mM IPTG and further incubation (at 18 °C for 20 hours or 37 °C for 3 hours).

Expressions were carried out in small scale (50 mL media). The pellets produced from these constructs were then subjected to test purifications, to see if Eaf5 would co-purify with the tagged Eaf3.

The results of these test purifications are inconclusive, Eaf3 may be observed close to the 50 kDa mark, though it is faint for both constructs, relative to contaminant bands. Conversely, Eaf5 (MW 31 kDa) is not observed by SDS-PAGE analysis, which should theoretically appear as intense as the tagged protein.

Biotin elutions of the sample are clear (data not included) and the protein can only be observed when 20 μL resin samples are loaded onto an SDS-PAGE gel. There may be a band present that corresponds to the C-terminal of Eaf5 (a common degradation product from other TINTIN purifications), but this coincides with denatured Strep-Tactin from the resin.

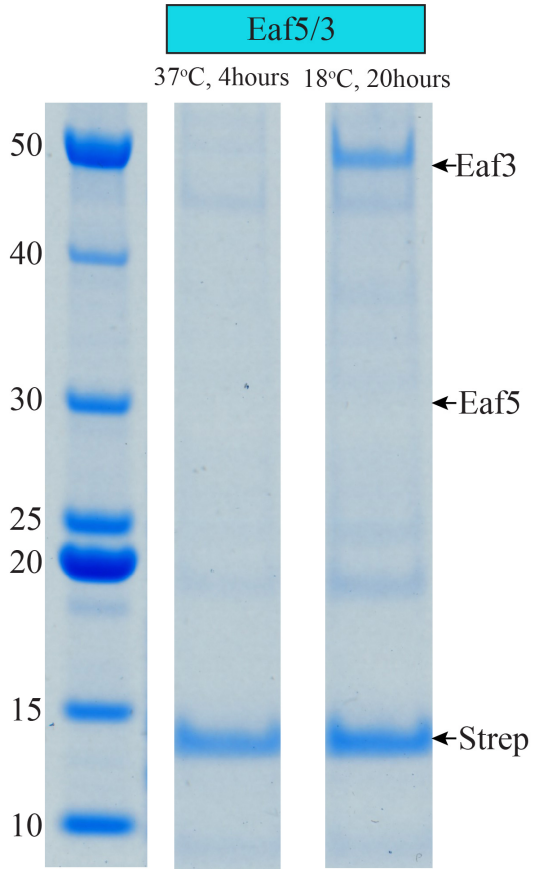

Figure 5.8: Small scale purifications (from 50 mL expressions) of an Eaf5/Eaf3 pET-28a construct. Containing a C-terminal Twin-Strep tag on the Eaf3 subunit. Two expression conditions are represented in two lanes. Each expression involved growing Rosetta 2 cells (transformed with the construct) at 37 °C , until OD reached ~0.6, before either being transferred to 18 °C for 20 hours, or further incubation at 37 °C for 4 hours. Samples are the result of loading 20 μL 1X SDS-PAGE loading buffer onto 50 μL Strep resin and boiling for 1 min. The potential position of subunit bands has been annotated, as well as the position of Strep-Tactin band from the resin.

5.4 Small-Angle X-ray Scattering

Small-angle X-ray scattering (SAXS) was carried out on two of the most prominent TINTIN constructs used throughout this project. This included the EA88A construct (Eaf7, 1-221) and EA88B (Eaf7, 1-260). Both constructs were purified as described in methods and results chapter 3.

EA88A was concentrated to 3.38 mg/mL and diluted in buffer A with two consecutive twofold dilutions to 1.51 mg/mL and 0.72 mg/mL. The EA88B construct was similarly diluted (3.59, 1.67 and 0.82 mg/mL), in order to generate a concentration series for SAXS analysis.

Data was collected at the BM29 BioSAXS beamline at ESRF Grenoble (under the supervision of Dr Nikos Pinotsis) and the following results were interpreted with the ATSAS software suite (as described in Methods).

5.4.1 Raw Data and Initial Analysis

Raw signal from each of the protein samples was averaged over 10 x 1 second frames with the buffer signal subtracted (**Figure 5.9**). The six resulting scattering data sets (2 constructs at 3 concentrations), were subjected to Guinier analysis with the Primus software package, to evaluate the quality of the data, as well as find an approximation for both the radius of gyration (R_g) and forward scattering intensity ($I(0)$) (**Figure 5.10**).

Guinier analysis shows no obvious signs of aggregation, with straight lines observed within the theoretical limits of the Guinier approximation. However, analysis of the Guinier plots (for both constructs EA88A and EA88B) showed increased estimations of R_g and $I(0)$, as the concentration of the the sample increased.

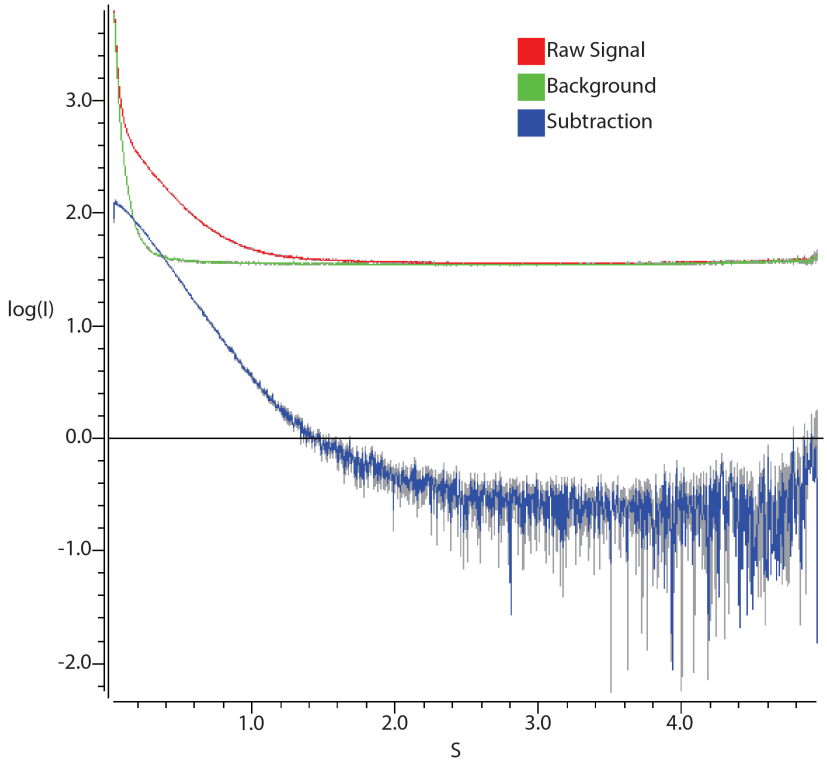

Figure 5.9: Representative raw data from SAXS analysis and subsequent buffer subtraction. Plot represents the log of intensity vs. scattering vector. Raw data, background buffer signal and buffer-subtracted signal, have been colour coded.

Raw data was further interpreted using PRIMUS/GNOM and subjected to indirect Fourier transform, to produce the paired distance distribution ($P(r)$) function (**Figure 5.11A and B**). *P(r)* distributions for both constructs produce asymmetric curves, suggesting that the protein contributing to the signal is in an extended conformation. D_{max} (the maximum dimension of the particle) was determined from these plots, as well as R_g, which was in agreement with Guinier estimations.

As with Guinier calculated values, the $I(0)$, R_g and D_{max} were found to increase with the concentration of the sample. For 88A, the D_{max} ranged from 20.5-26.6 nm (lowest to highest concentration), while 88B ranged from 19.6-26.51 nm. Kratky plots of the raw data (**Figure 5.11C and D**) suggests that both complexes contain some flexible regions. This extended/partially folded conformation was observed at all concentrations.

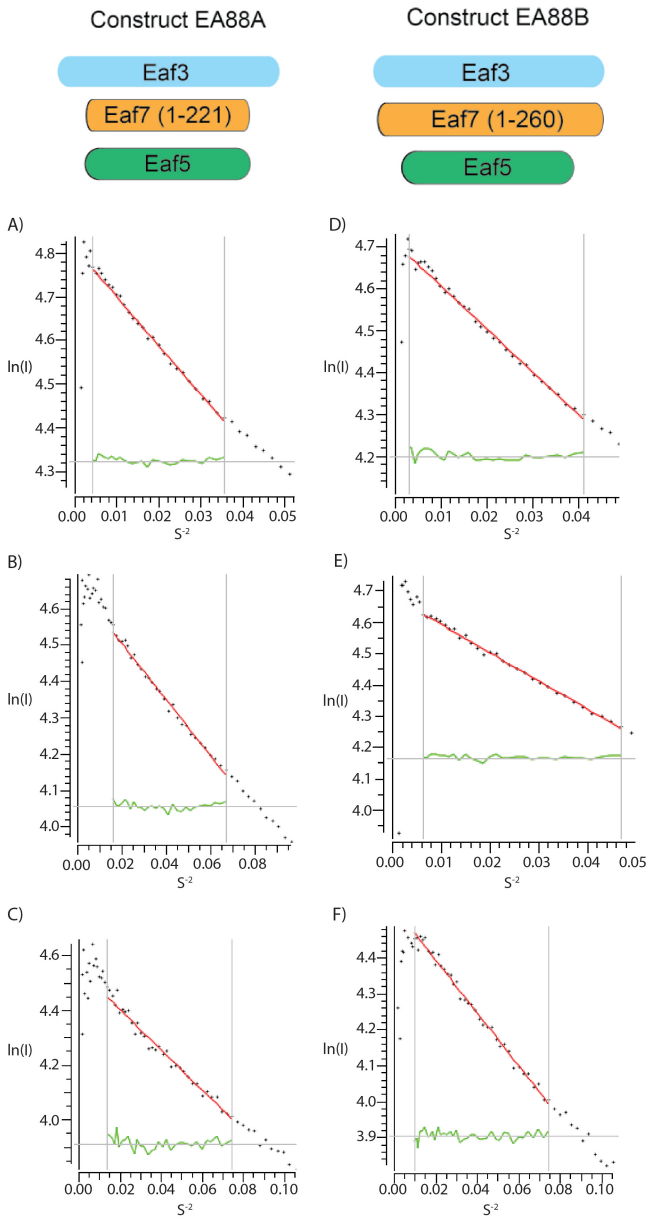

Figure 5.10: PRIMUS/GNOM autorg generated Guinier plots from SAXS raw data, acquired for both EA88A and EA88B truncated TINTIN constructs. Plotting the natural log of intensity vs the reciprocal of the scattering vector squared. Red lines indicate lines of best fit within the theoretical limits of the Guinier approximation and green lines display the residuals. **A–>C** displays Guinier plots of the EA88A construct of descending concentration (3.38–>0.72 mg/mL) and **D–>F** represents the EA88B construct (3.59–>0.82 mg/mL). For both constructs, extrapolation from the Guinier plot to determine R_g and $I(0)$, varies with the concentration of the sample.

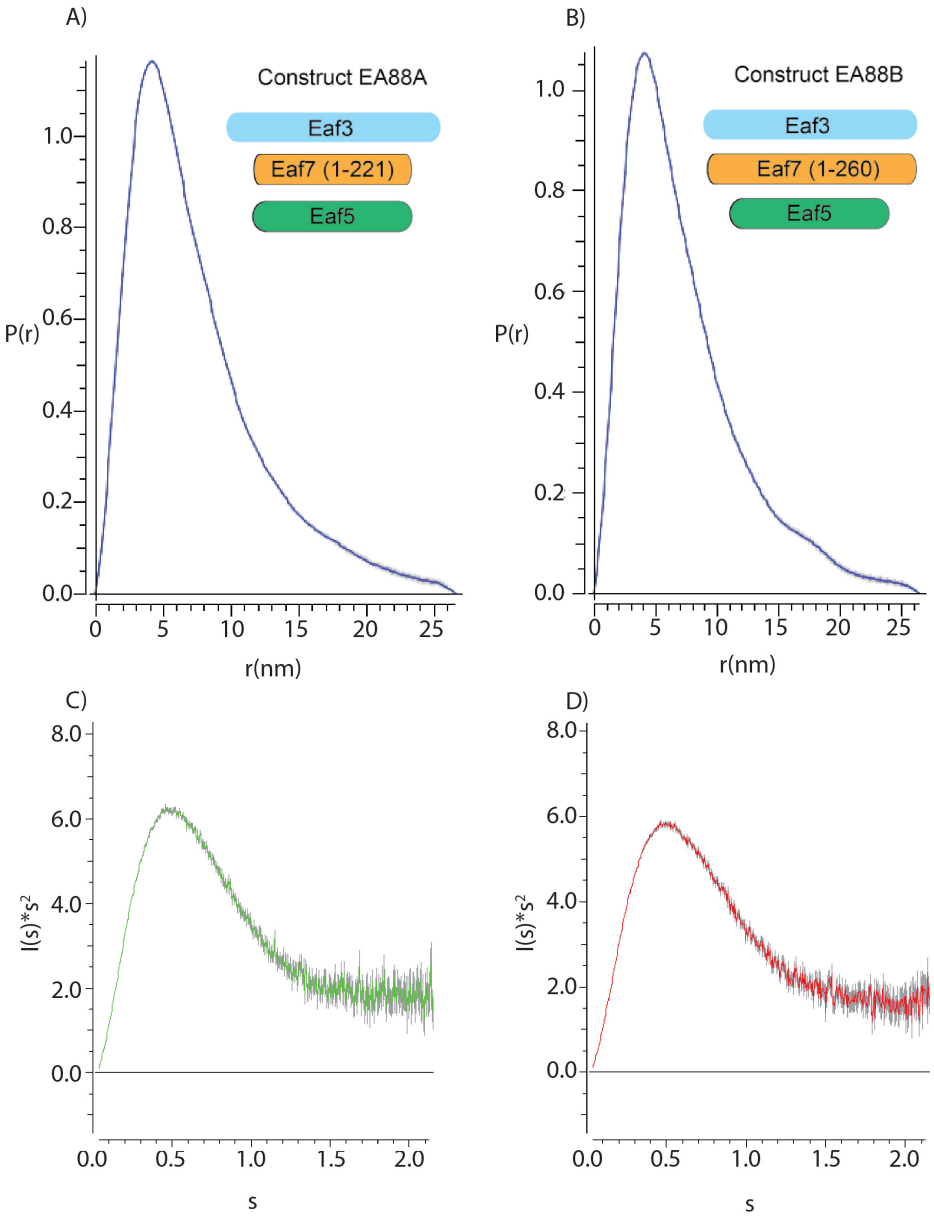

Figure 5.11: Representative P(r) function and Kratky analysis of both truncated TINTIN constructs. **A)** P(r) distribution of the EA88A complex. **B)** P(r) distribution of EA88B. Both P(r) functions show tails in their distribution (as opposed to a normal distribution) which suggests the complex is extended. **C and D)** Kratky plots of EA88A and EA88B respectively, suggesting that both complexes are in a folded state, though the tail at higher *s* values, suggests some flexibility within the complex.

	EA88A			EA88B		
Concentration (mg/mL)	3.38	1.51	0.72	3.59	1.67	0.82
I(0)	124	111	97.45	112.8	109	96
R_g (nm)	6.22	5.46	5.15	6.05	5.61	5.08
D_{max} (nm)	26.61	23.76	20.5	26.51	24.5	19.58
Porod Volume (nm^3)	262.6	230.9	221.4	249.9	228.0	214.0
Calculated MW(kDa)	115	99	89	104	101	88
Predicted MW (KDa)		106			111	

Table 5.1: Initial analysis from *P(r)* function of both the EA88A and EA88B constructs, comparing the calculated parameters at each concentration tested. Both sets of results were dependent on the concentration of the protein. MW was calculated using the I(0) of BSA standard (as described in methods).

BSA standards of known concentration, were used to calculate the MW of the species producing the scattering signal. Considering the variation in R_g and D_{max} with concentration, MW was calculated for each concentration independently (**Table 5.1.**). For EA88A (predicted MW: 106 kDa), the MW calculated ranged from 115 kDa (at highest concentration) to 89 kDa. Interestingly, the larger EA88B (predicted MW: 111 kDa), produced a smaller set of calculated molecular weight (104–>88 kDa). Calculating MW by Porod volume, resulted in higher estimations of MW compared to the $P(r)$ calculations, likely a reflection of noise at the higher values of s, in the raw data.

5.4.2 Initial Models

GNOM .out files from the highest and lowest concentration datasets, were processed by dammin to produce initial models (**Figures 5.12 and 5.13**). The $P(r)$ analysis was used to inform model building parameters, determining the use of a box shape appropriate for an elongated protein complex (cylindrical), with dimensions that would accommodate a structure with the calculated D_{max}. χ^2 values for the generated initial models (generated in a output .log file by dammin), were checked and found to be close to the ideal of ~1 and we were satisfied to continue processing.

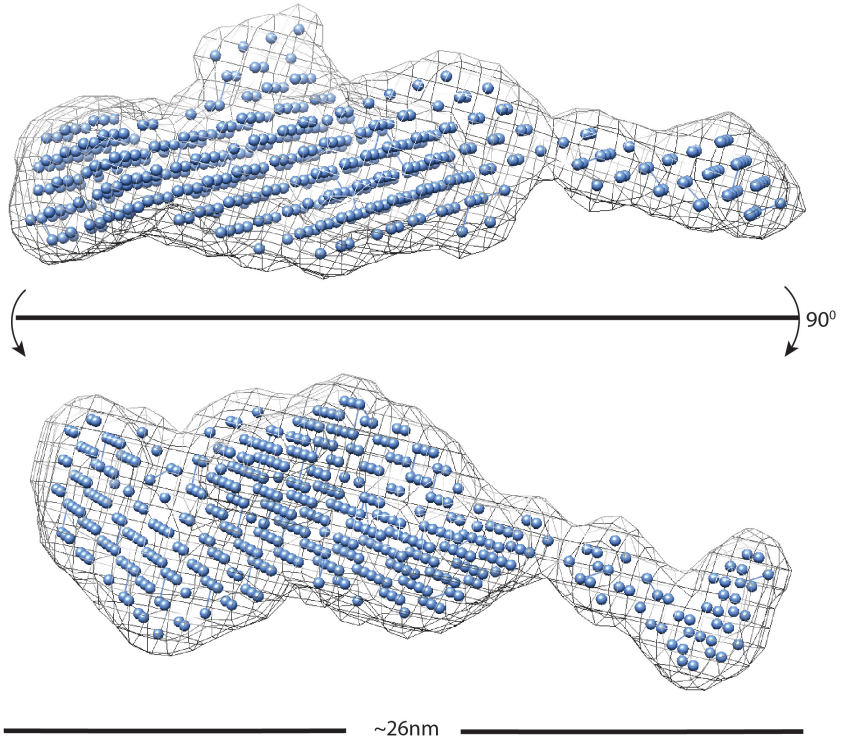

Figure 5.12: Representative dammin generated initial model of EA88A (Eaf7, 1-221) from the highest concentration dataset (3.38 mg/mL). The model is rotated 90° along the x-axis.

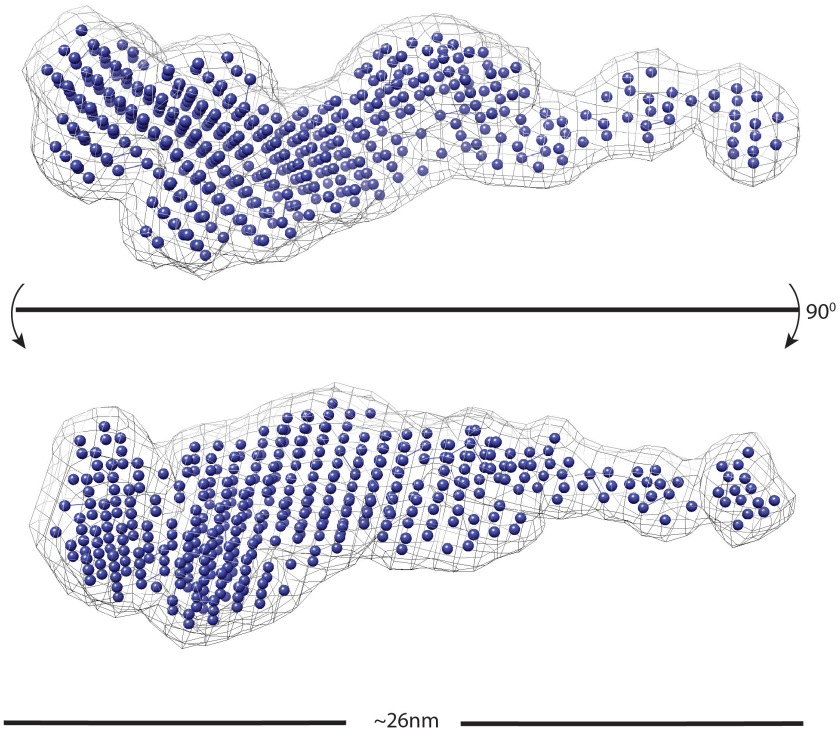

Figure 5.13: Representative of an initial model of EA88B (Eaf7, 1-260) produced by dammin. The model represents the highest concentration dataset (3.59 mg/mL). The model is rotated 90° along the x-axis.

Both complexes were purified separately and have a relatively small differences in predicted molecular weight (~5 kDa). Analysis of the raw data and initial models, suggest that both complexes are of similar MW (close to predicted MW of monomeric TINTIN, at higher concentrations) and dimensions.

5.4.3 Refining SAXS Models

Models were further processed using the damaver software suite (as described in methods). Damsel first aligned the low-resolution initial models using SUPCOMB and then generated a list of models, based on their normal spatial discrepancy values (NSD). Any model that deviated from the mean beyond the statistical constraints (NSD > 2 standard deviations + Mean NSD), were discarded, with the remaining models aligned by damsup (**Table 5.2**).

EA88A	Model	NSD (μ = 0.788)	Accept/Reject
	A	0.744	Accept
	B	0.745	Accept
	C	0.748	Accept
	D	0.750	Accept
	E	0.755	Accept
	F	0.768	Accept
	G	0.768	Accept
	H	0.777	Accept
	I	0.793	Accept
	J	0.814	Accept

Table 5.2: Representative summary of a damsel output file, listing all initial models produced by dammin and ordering them based on their normal spatial discrepancy. This particular dataset represents a TINTIN construct (EA88A, Eaf7 residues 1-221) at the highest concentration (3.38 mg/mL), with a mean NSD of 0.788 and an SD = 0.038. None of the models listed have an NSD > $(2xNSD + \mu)$, therefore all models were accepted and aligned with damsup.

Damsup aligned models were then averaged with damaver and filtered with damfilt (**Figures 5.14 and 5.15**). Comparison of high and low-concentration damfilt models, highlights the concentration dependence of the data, though the two TINTIN constructs share similar dimensions (**Figure 5.16 and 5.17**).

SAXS models of both constructs appear to be elongated, as the initial analysis of the raw data suggested. In both instances, the complex appears to be divided into a large lobe that extends into a narrower "tail" region. Higher concentration models appear larger than those produced from the lower concentration datasets, with the larger models ~5/6 nm longer than the lower concentration models.

However, both constructs produce low-resolution structures of similar shape and dimensions, with an approximate molecular weight close to theoretical MW of TINTIN (at the highest concentration values).

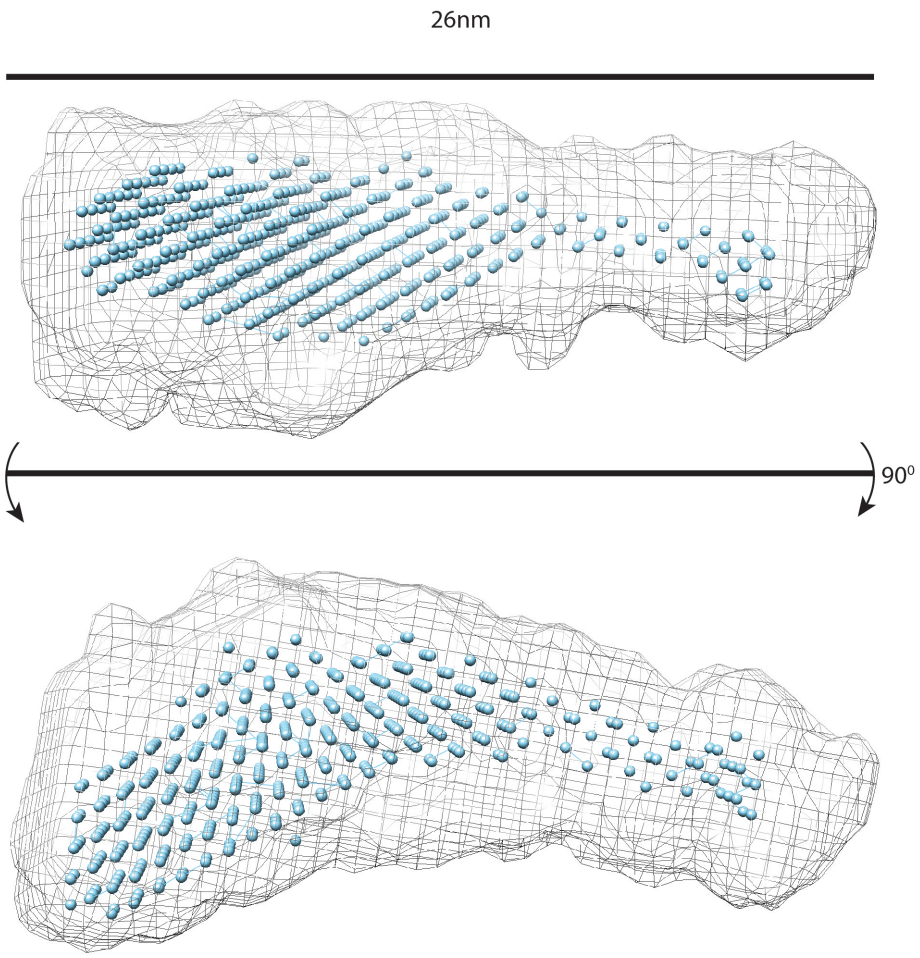

Figure 5.14: Damaver probability map and damfilt model of the TINTIN EA88A construct (Eaf7 residues 1-221, MW = 106 kDa). The structure was generated from the dataset, that corresponds to the highest concentration of the sample (3.38 mg/mL). The model is rotated 90° with respect to the x-axis, for a better representation of the entire structure. The filtered (damfilt output) model is represented in light blue, the surrounding mesh represents the damaver probability map.

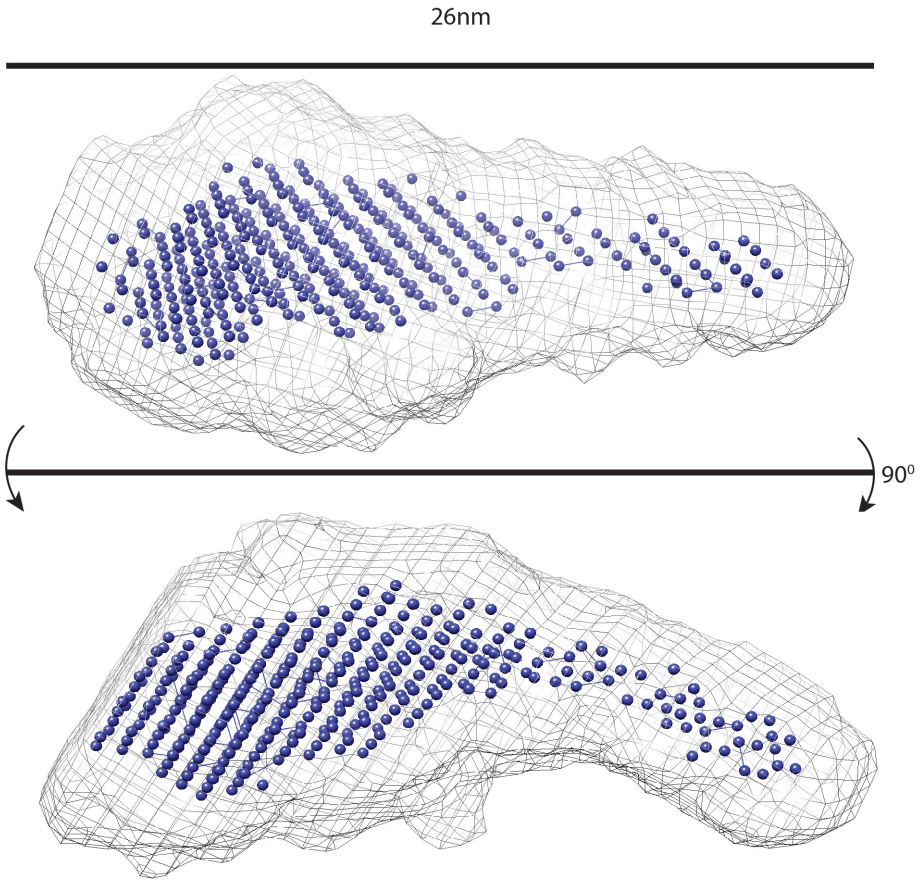

Figure 5.15: Damaver probability map and damfilt model of the TINTIN EA88B complex (Eaf7 residues 1-260, MW = 111 kDa). The structure was generated from the dataset that corresponds to the highest concentration of the sample (3.59 mg/mL). The model is rotated 90° with respect to the x-axis, for a better representation of the entire structure. The filtered (damfilt output) model is represented in dark blue, the surrounding mesh represents the damaver probability map.

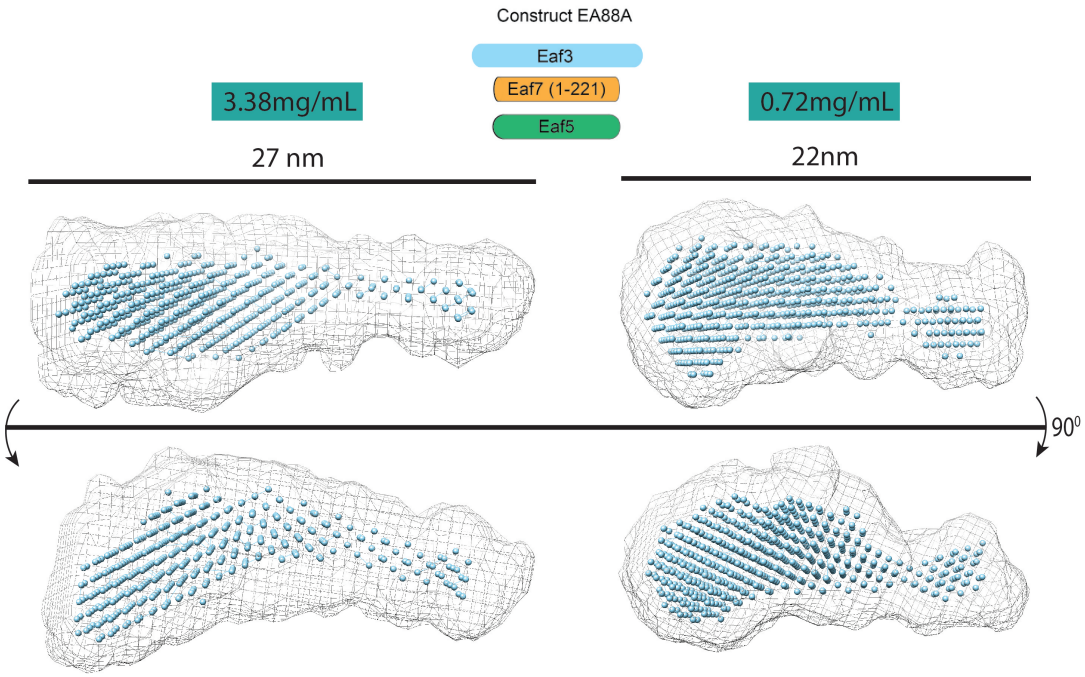

Figure 5.16: Comparison of EA88A TINTIN construct SAXS models, generated from highest and lowest concentration datasets(3.38 mg/mL and 0.72 mg/mL respectively). The damaver probability map is displayed as a mesh, with the damfilt average model displayed as light blue. The higher concentration model is approximately 5 nm longer than the lower concentration model, possibly the result of some oligomerization/ aggregation event. Both models show an extended conformation, containing a dense lobe that leads into a narrower tail region.

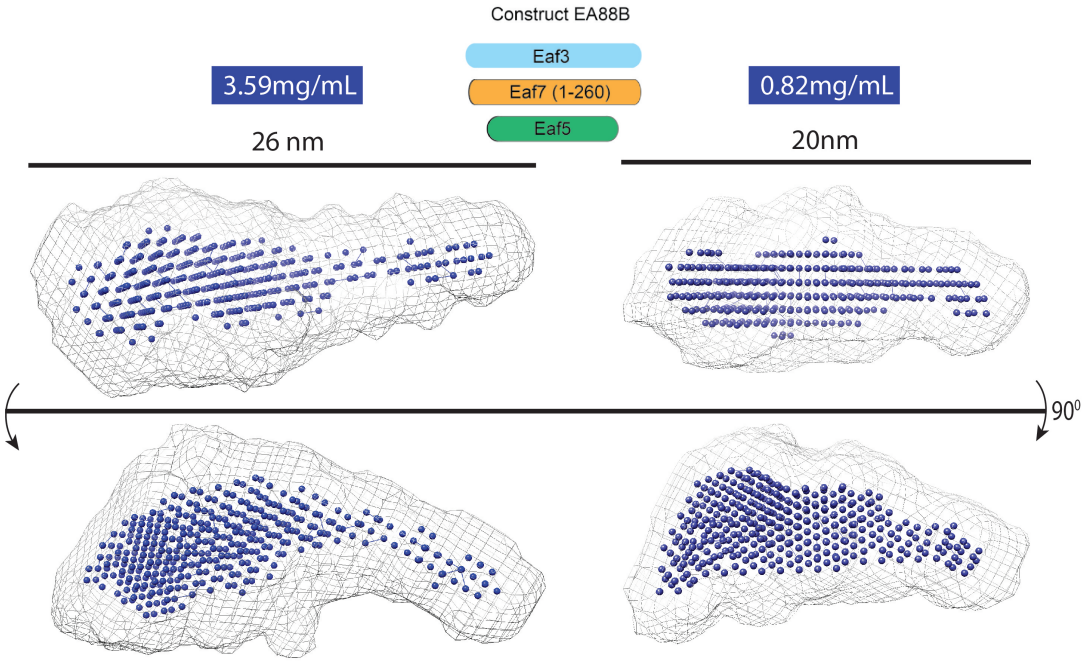

Figure 5.17: Comparison of EA88B TINTIN construct SAXS models, generated from highest and lowest concentration datasets (3.59 mg/mL and 0.82 mg/mL respectively). The damaver probability map is displayed as a mesh, with the damfilt average model displayed as dark blue. The higher concentration model is approximately 6 nm longer than the lower concentration model, possibly the result of some oligomerization/ aggregation event. Both models show an extended conformation, containing a dense lobe that leads into a narrower tail region.

Chapter 6

Architecture of the TINTIN Complex: Discussion

6.1 TINTIN Oligomeric State

SEC-MALS, IM-MS and SAXS analysis suggests that all purified TINTIN samples are dominated by the "monomeric" species (a single copy of Eaf7, Eaf3 and Eaf5). SEC-MALS chromatograms suggest that the most abundant species present in solution, is this monomeric species, but higher order aggregates are present in trace amounts for protein purified to <1 mg/mL. Ion mobility mass spectrometry analysis supports this finding, possibly revealing the presence of a dimeric TINTIN species, as well as Eaf5 and Eaf7 monomers. Other species that may correspond to TINTIN subunits, in complex with contaminants/degradation products, may also be present.

The high levels of noise and overlapping peaks in the native MS spectra, make it difficult to approximate relative abundance of each species accurately. Furthermore, we cannot rule out the possibility that the buffer exchange into ammonium acetate, followed by the vaporisation of sample by ESI, has altered the oligomeric composition of the purified mixture. Eaf7 and Eaf5 peaks for example, are not observed in SEC-MALS (or gel filtration purifications), neither is the suspected dimer of Eaf5/Eaf3, which raises questions as to whether these species are artefacts of the experimental procedure.

The Eaf5/Eaf3 dimer was an enticing prospect, as deletion of the Eaf5 subunit appeared to reduce the affinity between Eaf7 and Eaf3 (Chapter 3 and 4), which might hint at direct interactions between Eaf3 and Eaf5 that stabilise the complex. However, a bacterially expressed dimer has yet to be purified in preliminary trials and the feasibility of this dimer cannot be verified at present.

A current EM structure of the "core" TINTIN complex, proposes that Eaf5 forms a weak/transient interaction with the rest of the TINTIN complex (Wang et al. 2018), as Eaf3/Eaf7 were not present in the reconstruction.

CID of the TINTIN complex confirms the weaker/transient interaction between Eaf5 and the rest of the complex, as the subunit is shown to be ejected, while the subsequent Eaf3/Eaf7 dimer remains stable.

6.1.1 Unexpected Species

The 11 kDa species identified by IM-MS, may correspond to an Eaf5 degradation product, such as the contaminant species found to run close to the 15 kDa mark in SDS-PAGE gels. This contaminant is produced in all TINTIN purifications containing Eaf5 and was verified to be the C-terminal half of the protein, by the St. Andrews mass spectrometry service. This region has already been proposed to bridge TINTIN to the rest of NuA4 (Setiaputra et al. 2018).

Faint bands at the 10 kDa mark (SDS-PAGE analysis) were also noted to consistently appear in the final TINTIN purification product (Chapter 3), though these have yet to be identified.

If the 11 kDA species is an Eaf5 degradation product, it is potentially binding the Eaf3/Eaf7 dimer to form a truncated TINTIN complex, similar to the artificially generated constructs (EA125 and EA147). The 91 kDa species identified in the native spectra, approximates well to theoretical MW of the complex of truncated Eaf7, Eaf3 and the 11 kDa species. This may explain why the Eaf3/Eaf7 dimer was not identified in native spectra, despite the presence of Eaf5 monomer.

Weak interactions between Eaf5 and Eaf3/Eaf7 may result in an equilibrium between the complete complex and these component subunits. The 11 kDa species is potentially occupying the same interface as the full-length Eaf5 subunit, to form the 91 kDa species.

SAXS data may also support this hypothesis, as $P(r)$ function analysis of lower concentration datasets of the same TINTIN construct (EA88B), provides a calculated MW of ~88 kDa. However, the higher concentration dataset yields a calculated MW of 104 kDa. When the standard experimental error in MW calculations from SAXS data is considered ($\pm 10\%$), then the lower concentration dataset matches close to the 91 kDa species, whereas the higher concentration MW matches more closely to the TINTIN species containing full-length Eaf5.

Regardless as to whether this smaller TINTIN complex is represented by the 91 kDa species observed, it is clear that other species are present in the purified sample, though in trace amounts. These contaminants will certainly contribute to polydispersity in the sample and therefore offer a possible explanation, as to why TINTIN samples could not be crystallised.

6.1.2 TINTIN Structure

In the absence of protein crystals, high-resolution structures by X-ray crystallography of the purified complex, could not be produced. However, a combination of native IM-MS and SAXS data, provide insights into the overall dimensions of the EA88B TINTIN complex (Eaf7, 1-260).

Ion mobility data suggest that while individual TINTIN subunits (Eaf5 and truncated Eaf7), are extended/disordered to varying degrees, these species collapse into a more stable/compact conformation upon complex formation.

The CCS values produced from IM-MS experiments, weakly correlate to the square of R_g, that may be determined by other techniques such as SAXS (Chirot et al. 2010, Beveridge et al. 2014). The square root of calculated CCS values for the TINTIN complex is ~76 Å, which is significantly larger than the R_g calculations from SAXS data (~51 Å).

This could be a consequence of the weak relationship between CCS and R_g, or due to the fact that TINTIN has been exchanged into a different buffer and phase for IM-MS experiments, compared to SAXS experiments. CCS data also suggests that the TINTIN complex has CCS:mass and charge:mass ratios comparable to folded/ordered protein.

SAXS models from the highest and lowest concentration data points were developed for the EA88B complex (Eaf7, 1-260) and the smaller EA88A complex (Eaf7, 1-221). The fact that both complexes, with a ~5 kDa difference in MW, produce structures of similar shape and dimensions is encouraging. Truncated TINTIN is an extended complex of ~200-260 Å in length and ~60 Å diameter, which may explain the early elution profile in SEC purification steps (Chapter 3).

IM-MS data does not give indication of flexibility in the complex (though monomers and dimeric species indeed show signs of disorder). Conversely, Kratky plots from SAXS data, suggest that while the structure is folded, there may be regions of disorder/partial flexibility throughout the complex. Flexibility would also disrupt the process of crystal formation and when we also consider the presence of the contaminants previously discussed, it is not surprising that current crystallisation trials have failed.

The instability of the complex and the suspected flexibility observed, may potentially be overcome with the inclusion of additional subunits from the NuA4, Rpd3S and Pol II complexes. Interaction with other factors yet to be identified, may also contribute to increased stability. Further investigation is necessary to better evaluate TINTIN's biological function.

We attempted assigning the Eaf3 chromodomain and the partial Eaf5 structure (from the TEEAA cryo-EM structure) to a specific region within the SAXS models (data not included). Unfortunately, these higher resolution structures are significantly smaller than the overall SAXS model, which prevents accurate results from manual fitting, or more systematic assignment methods (using software such as CRYSOL).

Damfilt models were also manually fit into an existing negative stain structure of NuA4, which is proposed to contain the TINTIN complex, (Setiaputra et al. 2018). However, our SAXS structures do not fit particularly well within the negative-stain reconstruction (data not included). There could be several explanations for this discrepancy between SAXS models and existing models in the literature.

Firstly, the low-resolution negative stain structure of Setiaputra et al. (2018) may contain a partial TINTIN complex in their reconstruction (e.g Eaf5 but missing Eaf3 and/or Eaf7). This would not be unexpected, when we consider another (higher-resolution) structure of NuA4 (Wang et al. 2018) contains Eaf5 in the reconstruction, but not Eaf3 and Eaf7. Wang et al. speculate that Eaf5 creates a transient interaction with the rest of TINTIN, which is supported by our CID data.

Alternatively, given the extended conformation and the potential flexibility identified in Kratky analysis, it is possible that TINTIN undergoes a conformational change upon interaction with the rest of NuA4. There is precedent for this, as Wang et al. (2018) show that Eaf5 forms a highly ordered (α-helical) arrangement upon interaction with Tra1, while the native MS data reported within this investigation, shows that the free protein is highly disordered and folds upon incorporation into TINTIN. It is possible that, while TINTIN assembly results in correct folding of its component subunits, additional folding occurs, upon association with the rest of NuA4.

It is difficult to say at present, which of these two explanations is accurate, but high-resolution reconstructions of TINTIN, in complex with NuA4 and free of the coactivator complex, would provide a more definitive answer.

Chapter 7

Discussion and Future Directions

Summary

The following chapter takes into consideration all of the data presented herein and aims to frame it within the context of what is currently known about TINTIN and the parent NuA4 complex.

Firstly, the assembly of TINTIN subunits and the overall architecture of the subcomplex is discussed, drawing from truncation analysis. A summary figure is included, highlighting regions within the three subunits that are likely interacting, based on said truncation/test-purification analysis. The overall 3D structure of the complex is then discussed, using purification, native mass spectrometry and SAXS data, where we conclude TINTIN is likely extended, with some flexibility within the complex, though high-resolution structural data is not available at present. The oligomeric state of the complex is also discussed and while all biophysical/biochemical analysis indicates the presence of a single heterotrimer, higher order aggregates may be present in trace amounts, which in turn may be impeding the crystallisation process.

A discussion of TINTIN's interactions with Pol II CTD and chromatin, compares the pull-down data to what existing literature says about the overall role of the complex. TINTIN may indeed be interacting with both H3K36me3 and Pol II CTD, though the data cannot confirm the specificity or strength of these interactions. This data may be in agreement with the findings of Rossetto et al. (2014), who initially proposed a dual interaction between TINTIN and both H3K36me3 and phosphorylated CTD. However, the apparent weakness of binding observed in pull-down assays, may imply the need for additional factors in the TINTIN recruitment process. Moreover, due to the low sensitivity of these experiments, we cannot confirm an interaction with Pol II CTD, though increasing the length of the Eaf7 C-terminal (from residue 1-221 to 1-260), seems to facilitate low levels of binding.

7.1 TINTIN Assembly and Architecture

Considering its continued degradation throughout purifications, much of the Eaf7 C-terminal is likely disordered, in agreement with disorder prediction software. C-terminal truncations from residue 221, do not prevent association between the three subunits. However, C-terminal truncation from residue 142 does not clearly co-purify Eaf5 and may be weakening interactions between Eaf7 and Eaf3, despite the presence of the complete MBD (in Eaf7) and MRG domain (in Eaf3). This suggests an interaction between Eaf5 and residues 142-221 in Eaf7 and also implies points of contact between Eaf5 and Eaf3, that stabilise the complex. The latter interaction was not predicted in cross-linking mass spectrometry studies (Setiaputra et al. 2018), or pull-down experiments (Rossetto et al. 2014).

When Eaf7 N-terminal truncations (up to residue 86) are combined with C-terminal truncations (reducing Eaf7 to residues, 86-221 or 86-260), association with Eaf3 was disrupted. It should be noted that N-terminal truncations of Eaf7, remove half of the MBD, a domain that likely facilitates association with Eaf3's MRG domain, as observed in the orthologous MRG15/MRGBP complex (Xie et al. 2015).

However, this N-terminal truncation alone does not prevent Eaf3/Eaf7 association, as a complex is restored when Eaf7 C-terminal is present (up to residue 354). This may reveal additional points of contact made with Eaf3 with this region, that are sufficient to overcome the loss of a complete MBD in Eaf7.

Yeast Eaf7 MBD, may therefore be less critical for interactions with its partner Eaf3, compared to human MRG15/MRGBP complex. In addition to a lack of conservation of key residues, that were shown to make hydrophobic interactions between the two human domains (Xie et al. 2015), the 31-residue insertion in Eaf7's MBD, may contribute to weakening the interaction with Eaf3.

The yeast complex may utilise additional points of contact, beyond the boundaries of the MBD and MRG domain, perhaps relying on Eaf5 to stabilise the assembly.

Interaction between Eaf5 and Eaf7, is also prevented upon the truncation of the Eaf7 N-terminal, up to residue 86. Setiaputra et al. (2018), note the presence of a cross-link between Eaf7 residue 77 and Eaf5 residue 216. Both this data and the truncations reported herein, support interactions between Eaf7 N-terminal (between residues 1-86) and the Eaf5 C-terminal (residues 141-279), as essential for association between the two subunits. This is supported by our native mass spectrometry and purification data, which suggests that Eaf5 degradation products that have lost their N-terminal region, can still associate with the rest of the TINTIN complex.

Eaf5 N-terminal truncations, in combination with Eaf7 truncations (that did not disrupt complex assembly when full-length Eaf5 was expressed), also provide information on interacting regions. When Eaf5 is truncated from its N-terminal (Eaf5, 141-279), the truncated Eaf7 (Eaf7, 1-221) can no longer co-purify the rest of TINTIN. When Eaf7's C-terminal is extended to residue 260, the trimer is purified, though Eaf5 association may be reduced.

When we consider the purification results of Eaf5/Eaf7 truncations, compared to Eaf7 truncated constructs, we can identify boundaries for additional sites of interaction. The C-terminal of Eaf5 (141-279) may make contacts with Eaf7 between 221-260, as Eaf7 requires this region for interaction with the Eaf5 C-terminal. The N-terminal of Eaf5 (1-140) on the other hand, likely interacts with Eaf7 at residues 142-221, explaining why full-length Eaf5 can associate with Eaf7 (1-221), but not as strongly with the shorter Eaf7 (1-142).

Removal of Eaf5 entirely, prevents clear association between Eaf7 N-terminal and full-length Eaf3. Eaf7 (1-260), is capable of association with Eaf3 in the presence of full-length Eaf5 and the Eaf5 (141-279) construct, but complete removal of Eaf5 prevents this interaction.

The fact that this interaction is lost in the absence of Eaf5, further supports the notion that Eaf5 plays a stabilising role in TINTIN and may directly interact with Eaf3. This has yet to be predicted in other studies (Rossetto et al. 2014, Setiaputra et al. 2018, Wang et al. 2018).

The result of these purification attempts is contradictory to the CID results, which show that Eaf3/Eaf7 can dimerize (with the same C-terminal truncation in Eaf7, from residue 260). However, CID analysis is more sensitive to detecting labile complexes than by purification and SDS-PAGE. Considering the result of these mass spectrometry experiments, we cannot rule out the possibility that Eaf3 is indeed present in test purifications of truncated Eaf3/Eaf7 (Eaf7 1-260), though in lower abundance, due to the loss of Eaf5.

It is important to point out that an Eaf3/Eaf5 dimer could not be expressed/purified. While this does not rule out the proposed interactions between the two subunits, it is clear that any interaction is relatively weak and/or transient, which is why cross-linking experiments by Setiaputra et al. (2018) did not detect this interaction.

Earlier literature proposes that Eaf5 acts as an interface between TINTIN and the rest of NuA4, positioned on the periphery of the subcomplex, with a relatively weak affinity for Eaf3/Eaf7. This includes deletion/pull-down studies by Rossetto et al. (2014) and the cryo-EM structure of the TEEAA complex (Wang et al. 2018). The CID mass spectrometry data reported in this investigation, is indeed evidence that Eaf5 sits on the periphery of the TINTIN complex.

However, purifications, native MS, SAXS, SEC-MALS and truncation experiments, suggest that a stable/complete trimer is not only possible, but more abundant than any aggregates and dissociated subunits (at the concentrations tested).

This is in agreement with ChIP data from Rossetto et al. (2014), which shows Eaf5 localises with the rest of TINTIN, outside of the NuA4 complex. Eaf5 may therefore compensate for differences between Eaf3/Eaf7 and human MRG15/MRGBP subunits, that result in the weaker interactions we observe in the yeast subunits.

This stabilising role may prove to be one explanation for the presence of Eaf5 in yeast, with no known orthologs in higher Eukarya. Figure 7.1 represents a schematic of proposed TINTIN interaction interfaces, based on truncation studies and in agreement with existing literature. The figure highlights that points of contact, likely occur over relatively extensive boundaries.

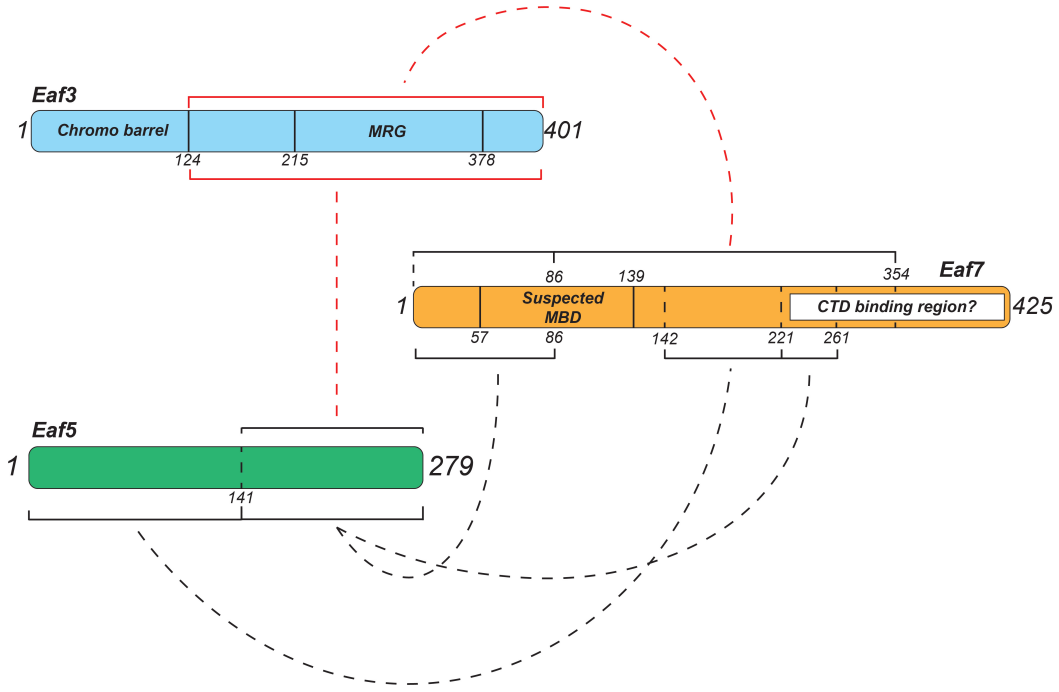

Figure 7.1: A schematic summary of inter-subunit interacting regions in the TINTIN subcomplex, deduced by truncation studies and in agreement with existing literature. Each schematic has been annotated to show the length of the primary sequence, as well as to identify the boundaries for conserved domains. Dashed lines connect surfaces proposed to interact, which are labelled with black bands and annotated with their residue boundaries. Eaf3 has been highlighted in red, to emphasise that this subunit has not been subjected to truncation analysis and exact boundaries for its interacting surfaces, cannot be deduced within the scope of this investigation.

SAXS data collected, particularly *P(r)* functions, indicate that TINTIN is in an extended conformation. The large interaction boundaries deduced by truncation analysis, supports the concept of an extended conformation. "Higher resolution" mutational analysis, coupled with structural data, may identify specific points of contact, but current data does suggest the potential for many points of contact along all three extended subunits.

In summary, Eaf7 association with Eaf5, is dependent on contact between the Eaf5 C-terminal (141-279) and two regions in Eaf7 (1-86 and 222-260). Additional points of contact may be made between Eaf5 N-terminal and residues 142-221 of Eaf7.

Eaf3/Eaf7 association is facilitated by a series of interactions between Eaf3 and residues 1-354 (in Eaf7). Loss of over 50% of the MBD in Eaf7 does not prevent association with Eaf3, provided full-length Eaf5 is present and the Eaf7 C-terminal includes residues up to position 354. The components essential for these interactions, distinguishes TINTIN from the human MRG15/MRGBP complex.

Future work would evaluate the essential regions within the Eaf3 subunit for TINTIN assembly and determine a minimal complex that may also allow for crystallisation. Interaction with additional factors (such as Pol II CTD, Eaf1 etc.) likely require a full-length complex (Rossetto et al. 2014) and efforts should be made to distinguish minimally stable truncated constructs, from fully functioning TINTIN. This will prove difficult *in vitro*, in the absence of catalytic activity, but may be measured with binding experiments such as pull-down assays or BLI.

7.2 Structure and Oligomeric State

Trace amounts of disordered monomeric species and higher order aggregates in the purified protein mixture, are likely the greatest detriment to the crystallisation of the TINTIN complex. While in low-abundance (based on purification and SEC-MALS data), future work would look to remove the source of this heterogeneity.

CID data from mass spectrometry analysis, suggests that the Eaf5 subunit forms a transient interaction with the rest of TINTIN, in agreement with Wang et al. (2018).

However, complete removal of Eaf5 has so far prevented the co-purification of Eaf3 via the Eaf7(1-260) subunit. Considering this data, questions regarding TINTIN stability *in vivo* must be raised, and whether Eaf3/Eaf7 form a dimer (as their human orthologs do) should be investigated further.

Native MS spectra obtained from the EA88B complex (Eaf7, 1-260), produced two species that have not been identified with confidence, at present. The 11 kDa and 91 kDa species in question, may correspond to the Eaf5 C-terminal (a degradation product in all TINTIN purifications) and a complex of this Eaf5 degradation product with the Eaf3/Eaf7 dimer, respectively. This could explain why Eaf3/Eaf7 dimer species were never observed in the native spectra, despite the presence of the corresponding Eaf5 monomer.

Moreover, truncation analysis confirms that the C-terminal of Eaf5, is capable of interacting with the rest of TINTIN (though contacts in the N-terminal also influences the strength of interaction). This is in agreement with cross-linking data from Setiaputra et al. (2018) and also supports the proposed identity of the two species detected in native MS.

Generating an Eaf3/Eaf7 dimer construct, could remove these contaminant species, assuming that they are indeed a consequence of Eaf5 degradation. However, constructs containing C-terminal truncated Eaf7, failed to co-purify Eaf3 (in the absence of Eaf5). Therefore, constructs containing a more complete Eaf7 must be produced and tested, for example, spanning residues 1-354.

Parallel to this, the 91 kDa species could be subjected to CID in repeat experiments. If the species does indeed correspond to a complex of Eaf7(1-260), Eaf3 and the Eaf5 C-terminal, then CID may yield an Eaf3/Eaf7 dimer and the 11 kDa species, similar to the ejection of Eaf5 that has already been observed.

The calculated dimensions of TINTIN constructs from SAXS analysis, was proportional to the concentration of the protein under investigation. The source of the concentration dependent effect on SAXS data is not clear, but diminishes the reliability of the final models produced. Removal of Eaf5 may simplify resulting datasets and comparison to trimeric datasets, may help determine the relative position of subunits within models.

Native MS and SEC-MALS analysis reveal trace amounts of high-MW contaminants, possibly aggregates of the TINTIN complex. These high-MW species and the presence of monomeric species (identified in native MS spectra), will certainly hinder SAXS investigations and the overall dimensions of the molecule cannot be quoted with certainty.

However, both truncated constructs EA88A (Eaf7, 1-221) and EA88B (Eaf7, 1-260), produce highly extended structures of similar dimensions, with molecular weights that approximate to that of a single TINTIN trimer (at the concentrations tested). The oligomeric state of the purified TINTIN species is further verified by native MS spectra.

These SAXS models do not fit well into the electron density assigned to TINTIN, within a NuA4 negative stain structure (Setiaputra et al. 2018). This may be due to the absence of Eaf3 and/or Eaf7 in the negative stain structure. Alternatively, this could be a consequence of additional folding/conformational changes, upon TINTIN interaction with the rest of NuA4.

For example, while our IM-MS data suggests Eaf5 is highly disordered, the protein adopts an ordered conformation in the presence of the rest of TINTIN. Eaf5 also adopts a more ordered, helical conformation when interacting with the Tra1 NuA4 subunit (Wang et al. 2018) and the TINTIN complex may adopt a more compact structure, upon interaction with the rest of NuA4.

While the exact dimensions calculated by SAXS analysis need further verification, initial analysis of *P(r)* functions and Kratky plots of both constructs at each concentration, agree on some key features of the complex. *P(r)* functions indicate the protein is extended, which may explain why the complex elutes from size-exclusion columns, at a volume that we might expect to find larger macromolecular assemblies. We note that Rossetto et al. (2014) observed a similar early elution, in their own size-exclusion purification protocol.

Kratky analysis indicates that while the complex contains regions of high order/stability, there may be regions of flexibility. This is again in agreement with our native ion mobility experiments, which confirm that the trimer has a CCS profile and charge state distribution indicative of an ordered complex. The partial structure of Eaf5 in the TEEAA complex, which depicts a subunit comprised of helices connected by loops/disordered regions, also support the results from this Kratky plot analysis.

Advances in electron microscopy, may lead to a high-resolution structural analysis of TINTIN, in the absence of successful crystallisation. While the full-length complex is relatively small, compared to complexes usually investigated with EM, mass spectrometry and SEC-MALS data, alludes to the presence of "dimeric" TINTIN at higher concentrations. This dimer may be suitable for the production of high-resolution EM structures. Whether higher order oligomers of TINTIN are biologically relevant, should be further explored, perhaps by native MS analysis of the endogenously expressed and purified complex.

7.3 TINTIN Interactions and Function

Due to weak interactions between TINTIN subunits and the modified H3 tail peptides, it is hard to discern differences in affinity, for post-translational modifications in pull-down experiments.

This also extends to potential interactions between TINTIN and Pol II CTD. The smaller of the TINTIN constructs (Eaf7, 1-221) has no apparent affinity for any of the CTD peptides, irrespective of the phosphorylation pattern, though the more complete of the two complexes (Eaf7, 1-260), may show a slightly increased affinity for the CTD.

On the other hand, based on current data, increasingly "complete" TINTIN constructs, do not appear to improve the affinity of Eaf3 for methylated histones, contrary to what was predicted at the beginning of this investigation. It may be that the complex offers no additional stabilisation for the chromodomain and is simply a molecular courier for the delivery of the Eaf3 subunit to the Rpd3S complex, after remodelling events carried out by the FACT chaperone complex (which Rossetto et al. (2014) showed directly interacts with TINTIN). The HDAC Rpd3S may in turn accommodate Eaf3, in a position that presents the chromodomain in a more favourable orientation, for interaction with methylated sites.

Work by Sun et al. (2008), also determined that the Eaf3 chromodomain had relatively weak affinity for H3K36me3, in agreement with our own pull-down data. The work of Rossetto et al. (2014), suggests that both interactions with methylated H3 and Pol II CTD, facilitate recruitment of TINTIN over coding regions. We could speculate that the weak affinity for both PTMs, necessitates this dual interaction for *in vivo* recruitment of TINTIN, at the correct position and time.

The fact that Eaf3 deletion does not completely abolish recruitment of Eaf7 to coding regions, implies that the strength of interaction between TINTIN and Pol II CTD, is sufficient for recruitment. Given the weak interaction observed in our *in vitro* pull-down assays, this seems unlikely. It is therefore more likely that additional factors facilitate TINTIN-Pol II interaction, which would explain why Rossetto et al. (2014) were able to isolate Pol II from cell-extracts using TINTIN, while our pull-down assays using purified products, did not show as strong an interaction.

The possibility of additional factors being necessary for recruitment, may be supported by a feature observed in the purification of the full-length TINTIN complex. The Eaf7 C-terminal rapidly degrades in all our preparations of full-length TINTIN, to the point where less than 50% of the subunit remains. This would be energetically wasteful in the *in vivo* context and may suggest the protein relies on additional factors, that not only stabilise, but assist in the recruitment process.

The low affinity for methylated H3 and Pol II CTD, may also enable a rapid turnover of TINTIN recruitment/release. If recruitment of TINTIN is dependent on the weak affinity for methylated H3 and phosphorylated CTD, this might ensure that Rpd3S recruitment of Eaf3 and subsequent demethylation activity, can only occur in the wake of Pol II (which is phosphorylated at Ser2 during the late elongation stage). In this way, TINTIN may be responsible for coordinating the activity of complexes such as FACT, Pol II and Rpd3S in the correct sequence, ensuring high rates of transcription (via FACT recruitment) and also the correct stabilisation of chromatin in the wake of Pol II (via Rpd3S recruitment).

A more sensitive, quantitative technique such as bio-layer interferometry (BLI) or isothermal titration calorimetry, would be better suited for determining subtle differences in affinity between protein and peptide. Such an investigation would also be able to determine whether or not there is a preference for specific histone (and Pol II CTD) PTMs. Increasingly "complete" TINTIN constructs, should also be evaluated for their peptide affinity, as we cannot dismiss the potential of a biological role for the Eaf7 C-terminal.

Considering Eaf5 may interact with Pol II CTD via phosphorylated Ser5 (Rossetto et al. 2014), future experiments would compare TINTIN containing full-length Eaf5 and the truncated variant, in their ability to pull-down Pol II. This would allow for the identification of Eaf5 components, that contribute to TINTIN localisation and activity.

The question as to whether TINTIN directly interacts with RNA polymerase II cannot be answered with certainty, due to the weak signal from pull-down assays. However, detection by SDS-PAGE is relatively insensitive, when compared to western blotting, from which the original interaction was proposed (Rossetto et al. 2014). Pull-down data may suggest that the EA88B construct (Eaf7, 1-260) is interacting weakly with Pol II CTD mimicking peptides, though considering the weak bands and no apparent difference between peptide affinity, we cannot dismiss a non-specific interaction.

However, this binding is not observed in the slightly smaller EA88A construct (Eaf7, 1-221), while sharing similar levels of binding for histone peptides. It may be that the C-terminal region of Eaf7, is essential for interaction between TINTIN and Pol II (**Figure 7.1**). Such a requirement would be unusual, as the C-terminal region is highly negatively charged, according to ProtParam predictions and evaluation of the primary sequence of the protein, which we would not expect to bind a highly phosphorylated sequence. Moreover, the C-terminal region of Eaf7 is not conserved with orthologs, such as MRGBP and we might expect some form sequence conservation, if this was the site of a specific interaction.

Ultimately, comparing the full-length TINTIN complex with these truncated constructs for their RNA Pol II affinity, would be highly informative, especially if the full-length Eaf7 shows higher affinity for Pol II CTD. If an interaction is indeed confirmed, then CTD peptides could be applied to purified TINTIN samples, in preparation for further crystallography trials, or even analysis by electron microscopy.

Future experiments would attempt to replicate the ChIP assays, carried out by Rossetto et al. (2014), using truncated TINTIN constructs generated in this investigation, that we observe to form a stable trimer. By doing this, we can attribute truncated regions to specific phenotypes, for instance the Eaf7 C-terminal truncations, may result in reduced recruitment over coding regions, if we are to assume this component interacts with Pol II CTD.

7.4 Final Remarks

While low-resolution models of the TINTIN complex have been produced, weak interactions between subunits, flexibility and trace quantities of higher order aggregates, have impeded the production of high-resolution structures. Comparison of SAXS models to partial structures and a negative stain structure of NuA4 in the existing literature, suggests that *in vitro* TINTIN is more extended, compared to when it associates with NuA4 (data not included). This may may reflect conformational changes upon association/dissociation with NuA4, or possibly be an artefact of the concentration dependence, observed in the initial SAXS data.

However, SAXS and native MS data does allude to an extended complex with regions of flexibility, which may explain the gel-filtration elution profile of the complex. Truncation analysis, suggests that points of contact are found across wide boundaries in the three subunits, which may prove problematic for efforts to remove these flexible regions, for the purposes of crystallisation.

TINTIN interacts with a number of complexes during transcription initiation through to elongation, which may include; Pol II, FACT, Rpd3S, NuA4, and SWR1. The proposed extended conformation of free TINTIN, may facilitate interactions with multiple complexes, by providing distinct binding interfaces.

CID data concurs with literature, that describes Eaf5 at the periphery of the complex, with relatively weak affinity for Eaf3/Eaf7. However, Eaf5 likely plays some additional stabilising role in TINTIN, compensating for the weak MBD-MRG interaction between Eaf7 and Eaf3, which is not as essential for maintaining subunit interaction (as observed in the human MRG15/MRGBP complex). The relatively weak association between Eaf3/Eaf7 (compared to the the human orthologous complex), could be the basis for Rpd3S recruitment and may enable the HDAC to outcompete TINTIN/NuA4, for the chromodomain containing Eaf3.

The delivery of Eaf3 by TINTIN to Rpd3S, may be a critical checkpoint after chromatin remodelling events, carried out by the FACT complex and in the wake of Pol II during the elongation process. This would place TINTIN as a conductor within the overall process, coupling the remodelling events necessary for high elongation rates, with the subsequent demethylation event, that stabilises chromatin after RNA polymerase II activity.

While pull-down assays potentially identify an interaction between Pol II CTD and a truncated TINTIN construct (Eaf7, 1-260), these results are inconclusive and will only be confirmed with more sensitive techniques.

The TINTIN complex resides at the interface between several key coactivator complexes and understanding its role within this sophisticated network of histone modification and transcription, requires further investigation. The features essential for complex assembly described herein, may inform further investigation, to identify functional elements within the complex and even supplement high-resolution structural analysis.

Bibliography

Aasland, R., Gibson, T. J. and Stewart, a. F. (1995) 'The PHD finger: implications for chromatin-mediated transcriptional regulation.', Trends in Biochemical Sciences, 20(2), pp. 56–9.

Allard, S. P. et al. (1999) 'NuA4, an essential transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the ATM-related cofactor Tra1p', The EMBO Journal, 18(18), pp. 5108–5119

Ameh, E. S. (2019) 'A review of basic crystallography and x-ray diffraction applications' The International Journal of Advanced Manufacturing Technology, 105, pp. 3289–3302

Auger, A. et al. (2008) 'Eaf1 is the platform for NuA4 molecular assembly that evolutionarily links chromatin acetylation to ATP-dependent exchange of histone H2A variants.', Molecular and Cellular Biology, 28(7), pp. 2257–70.

Avvakumov, N. and Côté, J. (2007) 'The MYST family of histone acetyltransferases and their intimate links to cancer.', Oncogene, 26(37), pp. 5395–5407.

Aylett, C. H. S. et al. (2016) 'Architecture of human mTOR complex 1', Science, 351(6268), pp. 48–52.

Baptista, T. et al. (2017) 'SAGA Is a General Cofactor for RNA Polymerase II Transcription', Molecular Cell, 68(1), pp. 130-143.

Baretić, D. and Williams, R. L. (2014) 'PIKKs - the solenoid nest where partners and kinases meet', Current Opinion in Structural Biology, 29, pp. 134–142.

Belotserkovskaya, R. et al. (2003) 'FACT Facilitates Transcription-Dependent Nucleosome Alteration', Science, 301(5636), pp. 1090–3.

Bennett, C. B. et al. (2001) 'Genes required for ionizing radiation resistance in yeast', Nature Genetics, 29(4), pp. 426–434.

Berndsen, C. E. et al. (2007) 'Nucleosome recognition by the Piccolo NuA4 histone acetyltransferase complex', Biochemistry, 46(8), pp. 2091–2099.

Bernecky, C. et al. (2016) 'Structure of transcribing mammalian RNA polymerase II', Nature, 529(7587), pp. 551–554.

Beveridge, R. et al. (2014) 'A mass-spectrometry-based framework to define the extent of disorder in proteins', Analytical Chemistry, 86(22), pp. 10979–10991.

Bhat, W., Ahmad, S. and Côté, J. (2015) 'TINTIN, at the interface of chromatin, transcription elongation, and mRNA processing.', RNA Biology, 12(5), pp. 486–9.

Bird, A. W. et al. (2002) 'Acetylation of histone H4 by Esa1 is required for DNA double-strand break repair', Nature, 419(September), pp. 411–415.

Biswas, D., Takahata, S. and Stillman, D. J. (2008) 'Different Genetic Functions for the Rpd3 (L) and Rpd3 (S) Complexes Suggest Competition between NuA4 and Rpd3 (S)', Molecular and Cellular Biology, 28(14), pp. 4445–4458.

Boudreault, A. A. et al. (2003) 'Yeast Enhancer of Polycomb defines global Esa1-dependent acetylation of chromatin', Genes and Development, 17(11), pp. 1415–1428.

Boyer, L. a, Latek, R. R. and Peterson, C. L. (2004) ‘The SANT domain: a unique histone-tail-binding module?’, Nature reviews. Molecular cell biology, 5(2), pp. 158–163.

Brown, C. E. et al. (2001) ‘Recruitment of HAT complexes by direct activator interactions with the ATM-related Tra1 subunit.’, Science, 292(June), pp. 2333–2337.

Cheung, A. C. M. and Diaz-Santin, L. M. (2019) ‘Share and share alike: the role of Tra1 from the SAGA and NuA4 coactivator complexes’, Transcription, 10(1), pp. 37–43.

Chirot et al. (2012) ’Statistical Analysis of Ion Mobility Spectrometry. Unbiased and Guided Replica-Exchange Molecular Dynamics’, American Society for Mass Spectrometry, 23, pp. 386-396.

Chittuluru, J. R. et al. (2011) ‘Structure and nucleosome interaction of the yeast NuA4 and Piccolo-NuA4 histone acetyltransferase complexes.’, Nature Structural & Molecular Biology, 18(11), pp. 1196–203.

Corden, J. L. (2016) ‘Pol II CTD Code Light’, Molecular Cell. Elsevier Inc., 61(2), pp. 183–184.

Côté, J. et al. (1994) ‘Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex.’, Science, 265(5168), pp. 53–60.

Diaz-Santin, L. M. et al. (2017) ‘Cryo-EM structure of the SAGA and NuA4 coactivator subunit Tra1 at 3.7 angstrom resolution’, eLife, 6, pp. 1–20.

Dobo, A. and Kaltashov, I. A. (2001) ‘Detection of multiple protein conformational ensembles in solution via deconvolution of charge-state distributions in ESI MS’, Analytical Chemistry, 73(20), pp. 4763–4773.

Doyon, Y. et al. (2004) ‘Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans.’, Molecular and Cellular Biology, 24(5), pp. 1884–96.

Doyon, Y. et al. (2006) ‘ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation’, Molecular Cell, 21(1), pp. 51–64.

Dubendorfft, J. W., Studier, W. and August, R. (1991) ‘Controlling Basal Expression in an Inducible T7 Expression’, Journal of Molecular Biology, 219, pp. 45–59.

Egelhofer, V. et al. (2002) ‘Protein identification by MALDI-TOF-MS peptide mapping: A new strategy’, Analytical Chemistry, 74(8), pp. 1760–1771.

Ehara, H. et al. (2017) ‘Structure of the complete elongation complex of RNA polymerase II with basal factors’, Science, 357(6354), pp. 921–924.

Eisen, A. et al. (2001) ‘The Yeast NuA4 and Drosophila MSL Complexes Contain Homologous Subunits Important for Transcription Regulation’, Journal of Biological Chemistry, 276(5), pp. 3484–3491.

Fenn, J. B. et al. (1989) ‘Electrospray ionization for mass spectrometry of large biomolecules’, Science, 246(4926), pp. 64–71.

Folta-Stogniew, E. (2006) ‘Oligomeric states of proteins determined by size-exclusion chromatography coupled with light scattering, absorbance, and refractive index detectors.’, Methods in Molecular Biology, 328, pp. 97–112.

Franke, D. et al. (2017) ‘ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions’, Journal of Applied Crystallography, 50, pp. 1212–1225.

Frost, B. et al. (2014) ‘Tau promotes neurodegeneration through global chromatin relaxation.’, Nature Neuroscience, 17(3), pp. 357–66.

Galarneau, L. et al. (2000) ‘Multiple Links between the NuA4 Histone Acetyltransferase Complex and Epigenetic Control of Transcription’, Molecular Cell, 5(6), pp. 927–937.

Garkavtsev, I. et al. (1996) ‘Suppression of the novel growth inhibitor p33ING1 promotes neoplastic transformation.’, Nature Genetics, 14(4), pp. 415–20.

Gavin, A.-C. et al. (2002) ‘Functional organization of the yeast proteome by systematic analysis of protein complexes.’, Nature, 415(6868), pp. 141–147.

Gibson, D. G. et al. (2009) ‘Enzymatic assembly of DNA molecules up to several hundred kilobases’, Nature Methods, 6(5), pp. 343–345.

Ginsburg, D. S., Govind, C. K. and Hinnebusch, A. G. (2009) ‘NuA4 Lysine Acetyltransferase Esa1 Is Targeted to Coding Regions and Stimulates Transcription Elongation with Gcn5’, Molecular and Cellular Biology, 29(24), pp. 6473–6487.

Govind, C. K. et al. (2010) ‘Phosphorylated Pol II CTD recruits multiple HDACs, including Rpd3C(S), for methylation-dependent deacetylation of ORF nucleosomes’, Molecular Cell, 39(2), pp. 234–246.

Gowher, H. et al. (2012) ‘Vezf1 protein binding sites genome-wide are associated with pausing of elongating RNA polymerase II’, Proceedings of the National Academy of Sciences, 109(7), pp. 2370–2375.

Grant, P. A. et al. (1997) ‘Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: Characterisation of an ada complex and the saga (spt/ada) complex’, Genes and Development, 11(13), pp. 1640–1650.

Guinier, A. (1969) ‘30 Years of Small-Angle X-Ray Scattering’, Physics Today, 22(11), pp. 25–30.

Gunduz, M., Ouchida, M. and Fukushima, K. (2000) ‘Genomic Structure of the Human ING1 Gene and Tumor-specific Mutations Detected in Head and Neck Squamous Cell Carcinomas Advances in Brief Detected in Head and Neck Squamous Cell Carcinomas 1’, Cancer Research, 3, pp. 3143–3146.

Hall, Z., Politis, A. and Robinson, C. V. (2012) ‘Structural modeling of heteromeric protein complexes from disassembly pathways and ion mobility-mass spectrometry’, Structure. Elsevier Ltd, 20(9), pp. 1596–1609.

Harper, T. M. and Taatjes, D. J. (2018) ‘The complex structure and function of Mediator’, Journal of Biological Chemistry, 293(36), pp. 13778–13785.

Heck, A. J. R. (2008) ‘Native mass spectrometry: A bridge between interactomics and structural biology’, Nature Methods, 5(11), pp. 927–933.

Heidemann, M. et al. (2013) ‘Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription’, Biochimica et Biophysica Acta - Gene Regulatory Mechanisms, 1829(1), pp. 55–62.

Hoke, S. M. T. et al. (2010) ‘Mutational analysis of the C-terminal FATC domain of Saccharomyces cerevisiae Tra1’, Current Genetics, 56(5), pp. 447–465.

Huang, J. and Tan, S. (2013) ‘Piccolo NuA4-catalyzed acetylation of nucleosomal histones: critical roles of an Esa1 Tudor/chromo barrel loop and an Epl1 enhancer of polycomb A (EPcA) basic region.’, Molecular and Cellular Biology, 33(1), pp. 159–69.

Huisinga, K. L., Brower-Toland, B. and Elgin, S. C. R. (2006) ‘The contradictory definitions of heterochromatin: Transcription and silencing’, Chromosoma, 115(2), pp. 110–122.

Ikura, T. et al. (2000) ‘Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis.’, Cell, 102(4), pp. 463–473.

Jacobs, S. A. et al. (2001) ‘Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3’, The EMBO Journal, 20(18), pp. 5232–5241.

Jacques, D. A. and Trewhella, J. (2010) ‘Small-angle scattering for structural biology — Expanding the frontier while avoiding the pitfalls’, Protein Science, 19(4), pp. 642–657.

Jenuwein, T. et al. (2001) ‘Translating the Histone Code’, Science, 293(5532), pp. 1074–1080.

Jiang, Y. W. and Stillman, D. J. (1996) ‘Epigenetic effects on yeast transcription caused by mutations in an actin-related protein present in the nucleus’, Genes and Development, 10(5), pp. 604–619.

Kelleher, R. J. Flanagan, P. M. and Kornberg, R. D. (1990) ‘A novel mediator between activator proteins and the RNA polymerase II transcription apparatus’, Cell, 61(7), pp. 1209–1215.

Knutson, B. A. and Hahn, S. (2011) ‘Domains of Tra1 Important for Activator Recruitment and Transcription Coactivator Functions of SAGA and NuA4 Complexes’, Molecular and Cellular Biology, 31(4), pp. 818–831.

Kirkwood, K. J. et al. (2013) ’Characterization of Native Protein Complexes and Protein Isoform Variation Using Size-fractionation-based Quantitative Proteomics’, Molecular Cell Proteomics, 12, pp. 3851-3873.

Konarev, P. V. et al. (2003) ‘PRIMUS: A Windows PC-based system for small-angle scattering data analysis’, Journal of Applied Crystallography, 36(5), pp. 1277–1282.

Kouzarides, T. (2007) ‘Chromatin Modifications and Their Function’, Cell, 128(4), pp. 693–705.

Kuo, M. H. et al. (1996) ‘Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific lysines’, Nature, 383, pp. 269–272.

Kurdistani, S. K. and Grunstein, M. (2003) ‘Histone acetylation and deacetylation in yeast.’, Nature Reviews. Molecular Cell Biology, 4(4), pp. 276–284.

Lai, W. K. M. and Pugh, B. F. (2017) ‘Understanding nucleosome dynamics and their links to gene expression and DNA replication’, Nature Reviews Molecular Cell Biology, 18(9), pp. 548–562.

Liu, G. et al. (2019) ‘Architecture of Saccharomyces cerevisiae SAGA complex’, Cell Discovery, 5(25), pp 1-4.

Loewith, R. et al. (2001) ‘Pho23 Is Associated with the Rpd3 Histone Deacetylase and Is Required for Its Normal Function in Regulation of Gene Expression and Silencing in Saccharomyces cerevisiae’, Journal of Biological Chemistry, 276(26), pp. 24068–24074.

Loidl, P. (1994) ‘Histone acetylation: facts and questions’, Chromosoma, 103(7), pp. 441–449.

Lorentz, A., Heim, L. and Schmidt, H. (1992) ‘The switching gene swi6 affects recombination and gene expression in the mating-type region of Schizosaccharomyces pombe’, MGG Molecular & General Genetics, 233(3), pp. 436–442.

Louder, R. K. et al. (2016) ‘Structure of promoter-bound TFIID and model of human pre-initiation complex assembly’, Nature, 531(7596), pp. 604–609.

Luger, K. et al. (1997) ‘Crystal structure of the nucleosome core particle at 2.8 A resolution.’, Nature, 389(6648), pp. 251–260.

Marklund, E. G. et al. (2015) ‘Collision cross sections for structural proteomics’, Structure. Elsevier Ltd, 23(4), pp. 791–799.

Le Masson, I. et al. (2003) ‘Yaf9, a novel NuA4 histone acetyltransferase subunit, is required for the cellular response to spindle stress in yeast.’, Molecular and Cellular Biology, 23(17), pp. 6086–102.

Maurer-Stroh, S. et al. (2003) ‘The Tudor domain “Royal Family”: Tudor, plant Agenet, Chromo, PWWP and MBT domains’, Trends in Biochemical Sciences, 28(2), pp. 69–74.

McMahon, S. B. et al. (1998) ‘The novel ATM-related protein TRRAP is an essential cofactor for the c- Myc and E2F oncoproteins’, Cell, 94(3), pp. 363–374.

McPherson, A. and Gavira, J. A. (2014) ‘Introduction to protein crystallisation’, Acta Crystallographica Section F:Structural Biology Communications. International Union of Crystallography, 70(1), pp. 2–20.

Mertens, H. D. T. and Svergun, D. I. (2010) ‘Structural characterization of proteins and complexes using small-angle X-ray solution scattering’, Journal of Structural Biology. Elsevier Inc., 172(1), pp. 128–141.

Mitchell, L. et al. (2008) ‘Functional dissection of the NuA4 histone acetyltransferase reveals its role as a genetic hub and that Eaf1 is essential for complex integrity.’, Molecular and Cellular Biology, 28(7), pp. 2244–56.

Narlikar, G. J., Sundaramoorthy, R. and Owen-Hughes, T. (2013) ‘Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes’, Cell, 154(3), pp. 490–503.

Natsume-Kitatani, Y., Shiga, M. and Mamitsuka, H. (2011) ‘Genome-wide integration on transcription factors, histone acetylation and gene expression reveals genes Co-regulated by histone modification patterns’, PLoS ONE, 6(7).

Neves, L. T. et al. (2017), 'The histone variant H2A.Z promotes efficient cotranscriptional splicing in S. cerevisiae', Genes and Development, 31(7), pp. 702-717

Nourani, A. et al. (2001) 'Role of an ING1 Growth Regulator in Transcriptional Activation and Targeted Histone Acetylation by the NuA4 Complex Role of an ING1 Growth Regulator in Transcriptional Activation and Targeted Histone Acetylation by the NuA4 Complex', Molecular and Cellular Biology, 21(22), pp. 7629–7640.

Olins, A. L. and Olins, D. E. (1974) 'Spheroid Chromatin Units (Bodies)', American Association for the Advancement of Science, 183(4122), pp. 330–332.

Piovesan, D. et al. (2018) 'MobiDB 3.0: More annotations for intrinsic disorder, conformational diversity and interactions in proteins', Nucleic Acids Research, 46(D1), pp. D471–D476.

Putnam et al. (2007) 'X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution', Quarterly Reviews of Biophysics, 40(3), pp. 191-285.

Reid, J. L., Moqtaderi, Z. and Struhl, K. (2004) 'Eaf3 Regulates the Global Pattern of Histone Acetylation in Saccharomyces cerevisiae', Molecular and Cellular Biology, 24(2), pp. 757–764.

Reines, D. (2003) 'Use of RNA Yeast Polymerase II Mutants in Studying Transcription Elongation', Methods in Enzymology, 371(1992), pp. 284–292.

Rosa, M. et al. (2013) 'The Arabidopsis SWR1 chromatin-remodeling complex is important for DNA repair, somatic recombination, and meiosis', Plant Cell, 25(6), pp. 1990–2001.

Rossetto, D. et al. (2014) 'Eaf5/7/3 form a functionally independent NuA4 submodule linked to RNA polymerase II-coupled nucleosome recycling.', The EMBO journal, 33(12), pp. 1397–415.

Saleh, A. et al. (1998) ‘Tra1p is a component of the yeast Ada.Spt transcriptional regulatory complexes’, Journal of Biological Chemistry, 273(41), pp. 26559–26565.

Santisteban, M. S., Kalashnikova, T. and Smith, M. M. (2000) ‘Histone H2A.Z regulates transcription and is partially redundant with nucleosome remodeling complexes.’, Cell, 103(3), pp. 411–422.

Schaft, D. et al. (2003) ‘The histone 3 lysine methyltransferase, SET2, is involved in transcriptional elongation’, Nucleic Acids Research, 31(10), pp. 2475–2482.

Schulze, J. M. et al. (2011) ‘Splitting the task: Ubp8 and ubp10 deubiquitinate different cellular pools of H2BK123’, Genes and Development, 25(21), pp. 2242–2247.

Selleck, W. et al. (2005) ‘The Saccharomyces cerevisiae Piccolo NuA4 histone acetyltransferase complex requires the Enhancer of Polycomb A domain and chromodomain to acetylate nucleosomes.’, Molecular and Cellular Biology, 25(13), pp. 5535–42.

Setiaputra, D. et al. (2018) ‘Molecular Architecture of the Essential Yeast Histone Acetyltransferase Complex NuA4 Redefines Its Multimodularity’, Molecular and Cellular Biology, 38(9), pp. 1–15.

Shimojo, H. et al. (2008) ‘Novel Structural and Functional Mode of a Knot Essential for RNA Binding Activity of the Esa1 Presumed Chromodomain’, Journal of Molecular Biology, 378(5), pp. 987–1001.

Shiv I. S. Grewall and Danesh Moazed (2003) ‘Heterochromatin and Epigenetic Control of Gene Expression’, Science, 301, pp. 798–802.

Sibanda, B. L. et al. (2017) ‘DNA-PKcs structure suggests an allosteric mechanism modulating DNA double-strand break repair’, Science, 355(6324), pp. 520–524.

Sivalingam, G. N et al. (2013) 'Amphirite: A program for processing travelling wave ion mobility mass spectrometry data', International Journal of Mass Spectrometry, 345-347, pp. 54-62.

Smith, E. R. et al. (1998) 'ESA1 is a histone acetyltransferase that is essential for growth in yeast.', Proceedings of the National Academy of Sciences of the United States of America, 95(7), pp. 3561–5.

Smolle, M. and Workman, J. L. (2013) 'Transcription-associated histone modifications and cryptic transcription', Biochimica et Biophysica Acta - Gene Regulatory Mechanisms, 1829(1), pp. 84–97.

Sterner, D. E. and Berger, S. L. (2000) 'Acetylation of histones and transcription-related factors.', Microbiology and Molecular Biology Reviews, 64(2), pp. 435–459.

Studier, F. W. (2005) 'Protein production by auto-induction in high-density shaking cultures', Protein Expression and Purification, 41(1), pp. 207–234.

Sun, B. et al. (2008) 'Molecular basis of the interaction of Saccharomyces cerevisiae Eaf3 chromo domain with methylated H3K36', Journal of Biological Chemistry, 283(52), pp. 36504–36512.

Svergun, D. I. (1992) 'Determination of the regularization parameter in indirect-transform methods using perceptual criteria', Journal of Applied Crystallography, 25(pt 4), pp. 495–503.

Svergun, D. I. (1999) 'Restoring low-resolution structure of biological macromolecules from solution scattering using simulated annealing', Biophysical Journal. Elsevier, 76(6), pp. 2879–2886.

Svergun, D. I and Koch, M. H. J. (2003) 'Small-angle scattering studies of biological macromolecules in solution', Reports on Progress in Physics, 66, pp. 147-227.

Tarazona, M. P. and Saiz, E. (2003) 'Combination of SEC/MALS experimental procedures and theoretical analysis for studying the solution properties of macromolecules', Journal of Biochemical and Biophysical Methods, 56(1–3), pp. 95–116.

Tong Ihn Lee and Richard A. Young (2000) 'Transcription of Eukaryotic Protein-Coding genes', Annual Review of Genetics, 341, pp. 77–137.

Tscherner, M. et al. (2012) 'The histone acetyltransferase Hat1 facilitates DNA damage repair and morphogenesis in Candida albicans', Molecular Microbiology, 86(5), pp. 1197–1214.

Uetrecht, C. et al. (2010) 'Ion mobility mass spectrometry of proteins and protein assemblies', Chemical Society Reviews, 39(5), pp. 1633–1655.

Uprety, B., Sen, R. and Bhaumik, S. R. (2015) 'Eaf1p Is Required for Recruitment of NuA4 in Targeting TFIID to the Promoters of the Ribosomal Protein Genes for Transcriptional Initiation In Vivo.', Molecular and Cellular Biology, 35(17), pp. 2947–64.

Utley, R. T. et al. (1998) 'Transcriptional activators direct histone acetyltransferase complexes to nucleosomes.', Nature, 394(July), pp. 498–502.

Volkov, V. V. and Svergun, D. I. (2003) 'Uniqueness of ab initio shape determination in small-angle scattering', Journal of Applied Crystallography. International Union of Crystallography, 36(3 I), pp. 860–864.

Wang, X. et al. (2018) 'Architecture of the Saccharomyces cerevisiae NuA4/TIP60 complex', Nature Communications, 9(1), pp. 1–12.

Wojnowska, M. et al. (2013) 'Autophosphorylation activity of a soluble hexameric histidine kinase correlates with the shift in protein conformational equilibrium', Chemistry and Biology, 20(11), pp. 1411–1420.

Wysocka, J. (2006) ‘Identifying novel proteins recognizing histone modifications using peptide pull-down assay’, Methods, 40(4), pp. 339–343.

Xie, T. et al. (2015) ‘Structural Basis for Multi-specificity of MRG Domains Article Structural Basis for Multi-specificity of MRG Domains’, Structure/Folding and Design, 23(6), pp. 1049–1057.

Xu, P. et al. (2016) ‘The NuA4 Core Complex Acetylates Nucleosomal Histone H4 through a Double Recognition Mechanism’, Molecular Cell, 63(6), pp. 965–975.

Yamamoto, T. and Horikoshi, M. (1997) ‘Novel Substrate Specificity of the Histone Acetyltransferase Activity of HIV-1-Tat Interactive’, The Journal of Biological Chemistry, 272(49), pp. 30595–30598.

Yan, Y. et al. (2000) ‘Crystal structure of yeast Esa1 suggests a unified mechanism for catalysis and substrate binding by histone acetyltransferases.’, Molecular cell, 6(5), pp. 1195–1205.

Yang, H. et al. (2013) ‘MTOR kinase structure, mechanism and regulation’, Nature, 497(7448), pp. 217–223.

Yuan, H. et al. (2012) ‘MYST protein acetyltransferase activity requires active site lysine autoacetylation.’, The EMBO Journal, 31(1), pp. 58–70.

Zhang, B. et al. (2014) ‘Histone phosphorylation: Its role during cell cycle and centromere identity in plants’, Cytogenetic and Genome Research, 143(1–3), pp. 144–149.

Zhang, H. et al. (2004) ‘The Yaf9 component of the SWR1 and NuA4 complexes is required for proper gene expression, histone H4 acetylation, and Htz1 replacement near telomeres.’, Molecular and Cellular Biology, 24(21), pp. 9424–9436.

Zhang, P. et al. (2006) 'Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3', Nucleic Acids Research, 34(22), pp. 6621–6628

Zhang, Y. and Reinberg, D. (2001) 'Transcription regulation by histone methylation: Interplay between different covalent modifications of the core histone tails', Genes and Development, 15(18), pp. 2343–2360.

Zhao, K. et al. (1998) 'Rapid and Phosphoinositol-Dependent Binding of the SWI/SNF-like BAF Complex to Chromatin after T Lymphocyte Receptor Signaling', Cell, 95(5), pp. 625–636.

Zhu, J. et al. (2015) 'Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth.', Nature, 525(7568), pp. 206–11.

Appendix

9.1 Appendix A

Prediction	*Eaf7*	*Eaf5*	*Eaf3*
MobiDB-lite	**66.6%**	**11.8%**	**30.2%**
DisEMBL-465	57.4%	11.1%	32.9%
DisEMBL-HL	74.1%	39.8%	44.6%
ESpritz-DisProt	54.4%	4.3%	54.4%
Espritz-NMR	82.2%	28.7%	48.6%
Espritz-XRay	74.6%	16.1%	42.4%
GlobPlot	30.1%	15.4%	25.9%
IUPred-long	76.2%	17.2%	33.4%
IUPred-short	71.8%	16.8%	32.9%
JRONN	76.0%	25.1%	45.1%
VSL2b	76.9%	45.2%	47.9%

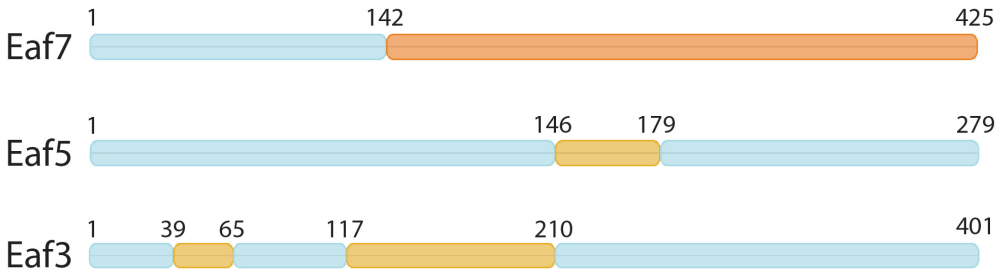

Figure 9.1: MobiDB disorder prediction results for the full-length TINTIN subunits. The predicted percentage of disorder in each subunit is listed for each corresponding prediction software result. Schematics represent the mobiDB-lite averaged disorder predictions, and highlights regions of disorder in yellow or orange. Eaf7 is predicted to be highly disordered between residue 142 and the C-terminal. Eaf5 is predicted to be primarily ordered, with the exception of a small region of disorder (146-179). Eaf3 contains two regions of disorder in its N-terminal half, while its C-terminal is predicted to be ordered.

9.2 Appendix B

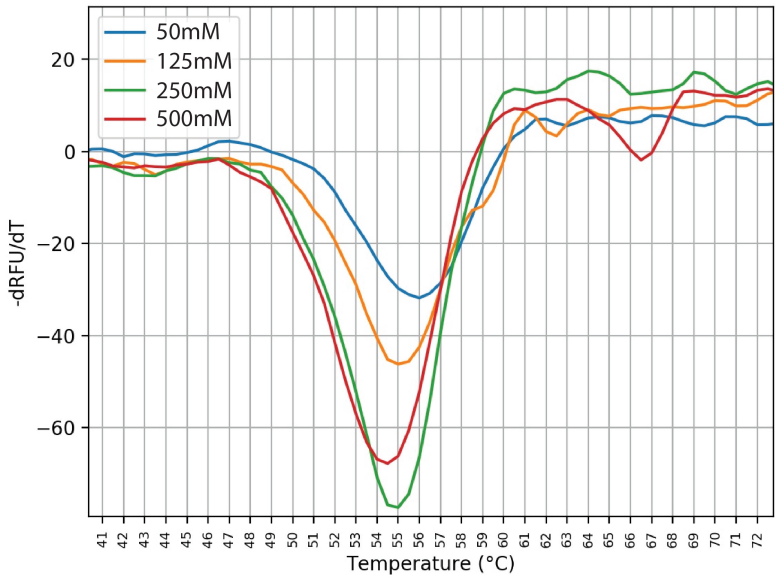

Figure 9.2: Example of thermofluor dataset of the full-length TINTIN construct, in HEPES buffer (pH 7.5) in increasing concentrations of NaCl. Plot is the change in fluorescence over time vs temperature. The T_m is defined as the temperature at which half of the protein species in solution is denatured. Comparison on the x values of minima of each trace (which represents TINTIN in a different buffer condition) compares their T_m. More stabilising buffers will produce higher T_m values.

	T_m values (°C)		
	100 mM SPG	100 mM Hepes	100 mM Tris.HCl
50 mM	56.5	56.0	56.0
125 mM	55.5	55.0	55.0
250 mM	55.5	54.7	54.5
500 mM	55.0	54.5	54.0

Table 9.1: Summary of thermofluor experimental results, where 2.7 mg/mL of full-length TINTIN is incubated in varying concentrations of salt/buffer from the RUBIC commercial screen. All buffers listed here were used at the same concentration and pH (100 mM, pH 7.5). T_m values listed above are the average from duplicate experiments. Increasing concentrations of salt appear to reduce TINTIN stability, irrespective of buffer used.

9.3 Appendix C

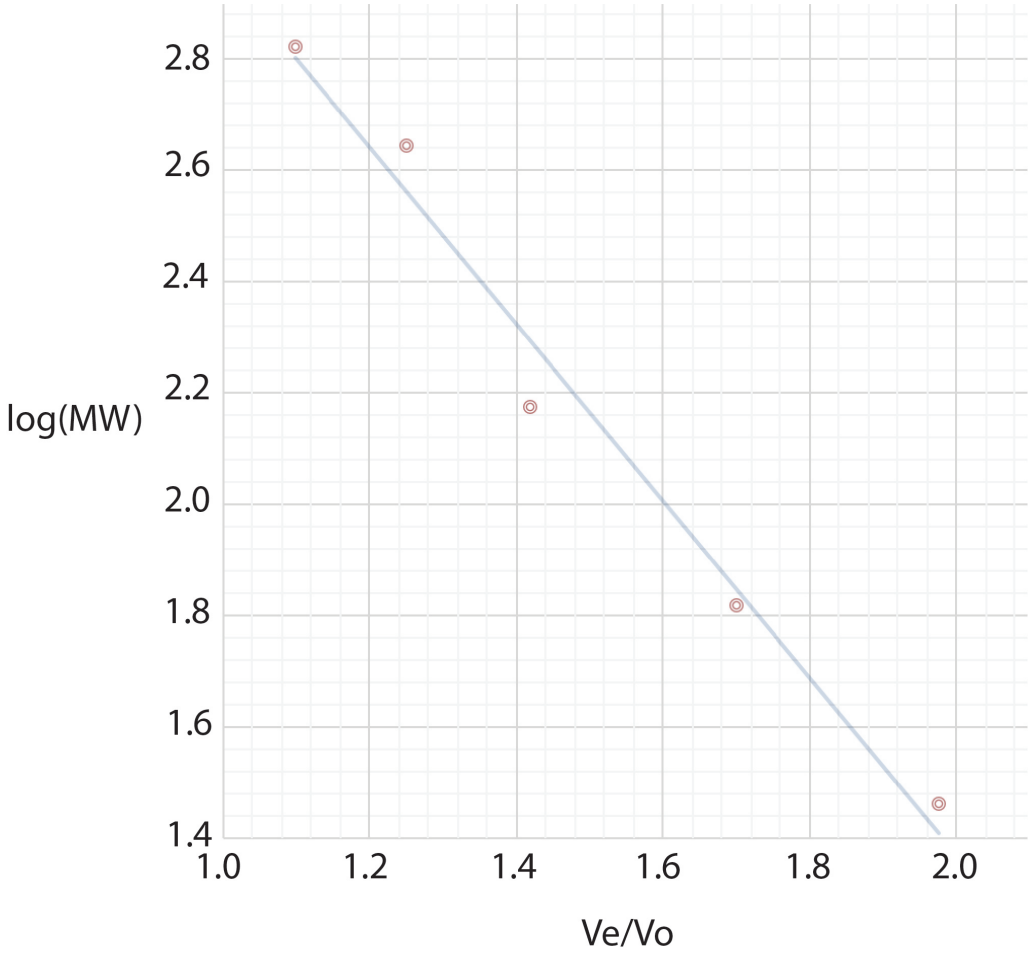

Figure 9.3: Calibration curve of the HiLoad Superdex 200 PG 16/600 column, courtesy of Dr Alan Cheung. 5 protein standards of molecular weight ranging from 29-669 kDa were eluted through the column. Elution volumes (V_e) are divided by the void volume (V_o) which is determined by the elution of a large molecule such as blue-dextran. Plotting the log of the MW of these species against V_e/V_o provides a line of best fit that can be used to approximate the MW of experimental molecules during purification.